

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 14:59:04 ; Search time 338 Seconds
(without alignments)
9462.933 Million cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049
Sequence: 1 ttcaagtgctactactatgg.....ttgaagtgctactactatg 2049

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 1085931 seqs, 780495707 residues

Word size: 20

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database: Published Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEM_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by change to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	29	1.4	1194	10	US-09-739-907-49
2	29	1.4	1222	10	US-09-739-907-31
3	22	1.1	4022	9	US-09-964-899-18
4	20	1.0	504	9	US-09-918-995-20201

ALIGNMENTS

RESULT 1
US-09-739-907-49
Sequence 49, Application US/09739907
Patent No. US20010012889A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 36 Human Secreted Proteins
FILE REFERENCE: P2022P1
CURRENT APPLICATION NUMBER: US/09/739, 907
CURRENT FILING DATE: 2000-12-20

PRIOR APPLICATION NUMBER: 09/348,457
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: 60/070,567
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,692
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,704
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,658
PRIOR FILING DATE: 1998-01-07
NUMBER OF SEQ ID NOS: 196
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 49
LENGTH: 1194
TYPE: DNA
ORGANISM: Homo sapiens
US-09-739-907-49

Query Match 1.4%; Score 29; DB 10; Length 1194;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 CAGATATCAATTGATGGCTATGCAAA 81
|||||
Db 478 CAGATATCAATTGATGGCTATGCAAA 506

RESULT 2
US-09-739-907-31
Sequence 31, Application US/09739907
Patent No. US20010012889A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 36 Human Secreted Proteins
FILE REFERENCE: P2022P1
CURRENT APPLICATION NUMBER: US/09/739, 907
CURRENT FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 09/348,457
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: 60/070,567
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,692
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,704
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,658
PRIOR FILING DATE: 1998-01-07
NUMBER OF SEQ ID NOS: 196
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 31
LENGTH: 1222
TYPE: DNA
ORGANISM: Homo sapiens
US-09-739-907-31

Query Match 1.4%; Score 29; DB 10; Length 1222;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 CAGATATCAATTGATGGCTATGCAAA 81
|||||
Db 501 CAGATATCAATTGATGGCTATGCAAA 529

RESULT 3
US-09-964-899-18
Sequence 18, Application US/09964899
Patent No. US20020174446A1
GENERAL INFORMATION:
APPLICANT: Cohen, Dalja et al.
TITLE OF INVENTION: Identification of Genes Involved in
Alzheimer's Disease Using Drosophila Melanogaster
FILE REFERENCE: 4-31612 A

; CURRENT APPLICATION NUMBER: US/09/964,899
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,893
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/298,309
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 18
; LENGTH: 4022
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(4022)
; OTHER INFORMATION: n - A,T,C or G
US-09-964-899-18

Query Match 1.1%; Score 22; DB 9; Length 4022;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 922 CAGAGGCCAGGAGCTGTATG 943
DB 319 CAGAGGCCAGGAGCTGTATG 340

RESULT 4
US-09-918-995-20201/c
; Sequence 20201, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 20201
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(504)
; OTHER INFORMATION: n - A,T,C or G
US-09-918-995-20201

Query Match 1.0%; Score 20; DB 9; Length 504;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1056 TGCCCTGGGCTACAGTTTG 1075
DB 78 TGCCCTGGGCTACAGTTTG 59

Search completed: July 5, 2003, 17:40:52
Job time : 339 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 12:55:04 ; Search time 100 Seconds
(without alignments)
6283.807 Million cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttcaagtgtactactctg.....ttgaagtgtactacttag 2049

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 20

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents-NA:*

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PTCUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Dackfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	1.1	4020	4	US-09-050-159-130
2	22	1.1	4024	4	US-09-162-484-18
3	21	1.0	3942	4	US-09-162-484-19

ALIGNMENTS

RESULT 1
US-09-050-159-130
; Sequence 130, Application US/09050159A
; Patent No. 6197505
; GENERAL INFORMATION:
; APPLICANT: No. 6197505berg, Leif T
; APPLICANT: Andersson, Maria K
; APPLICANT: Linstrom, Per H
; TITLE OF INVENTION: METHODS FOR ASSESSING CARDIOVASCULAR STATUS AND
; FILE REFERENCE: 1248/ID042
; CURRENT APPLICATION NUMBER: US/09/050,159A
; CURRENT FILING DATE: 1998-03-27
; EARLIER APPLICATION NUMBER: 60/042,930
; EARLIER FILING DATE: 1987-04-03
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 130
; LENGTH: 4020

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Angiotensin I converting enzyme mRNA
US-09-050-159-130

Query Match 1.1%; Score 22; DB 4; Length 4020;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 922 CAGAAGCCCAAGAGCTGTATG 943
DB 317 CAGAAGCCCAAGAGCTGTATG 338

RESULT 2
US-09-162-484-18
; Sequence 18, Application US/09162484
; Patent No. 6248724
; GENERAL INFORMATION:
; APPLICANT: Phillips, M. Ian
; APPLICANT: Mohuczy, Dagmar
; TITLE OF INVENTION: ANTI-SENSE OLIGONUCLEOTIDE COMPOSITIONS TARGETED TO
; FILE REFERENCE: UFLA:087/UFLA087P
; CURRENT APPLICATION NUMBER: US/09/162,484
; CURRENT FILING DATE: 1998-09-25
; EARLIER APPLICATION NUMBER: 60/059,661
; EARLIER FILING DATE: 1997-09-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 4024
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-162-484-18

Query Match 1.1%; Score 22; DB 4; Length 4024;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 922 CAGAAGCCCAAGAGCTGTATG 943
DB 317 CAGAAGCCCAAGAGCTGTATG 338

RESULT 3
US-09-162-484-19
; Sequence 19, Application US/09162484
; Patent No. 6248724
; GENERAL INFORMATION:
; APPLICANT: Phillips, M. Ian
; APPLICANT: Mohuczy, Dagmar
; TITLE OF INVENTION: ANTI-SENSE OLIGONUCLEOTIDE COMPOSITIONS TARGETED TO
; FILE REFERENCE: UFLA:087/UFLA087P
; CURRENT APPLICATION NUMBER: US/09/162,484
; CURRENT FILING DATE: 1998-09-25
; EARLIER APPLICATION NUMBER: 60/059,661
; EARLIER FILING DATE: 1997-09-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 3942
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-162-484-19

Query Match 1.0%; Score 21; DB 4; Length 3942;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 923 AGAAGCCCAAGAGCTGTATG 943

Mon Jul 7 09:49:08 2003

us-09-937-521-15.oli20.rni

Page 2

Db 314 |||||
AGAGGCCAAGAGCTGTATG 334

Search completed: July 5, 2003, 15:50:30
Job time : 100 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 12:49:14 ; Search time 2980 Seconds
(without alignments)
11135.756 Million cell updates/sec

Title: US-09-937-521-15
Perfect score: 2049
Sequence: 1 ttcatgtgcactactatgg.....ttgaagtgtcactactatg 2049

Scoring table: ORIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues
Word size : 20

Total number of hits satisfying chosen parameters: 44

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST: *
1: em_estba:*
2: em_esthum:*
3: em_estnu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	655	32.0	700	10	BB356080 BB356080
2	536	26.2	896	12	BP237061 602027402
3	157	7.7	451	9	AA920146 vY2h04.1
4	155	7.6	674	17	AZ018768 RPCI-23-3
5	125	6.1	834	17	BH049419 RPCI-24-3
6	112	5.5	269	9	AV284970 AV284970

7	100	4.9	292	17	AZ907075
8	97	4.7	451	17	AZ879271 RPCI-23-1
9	85	4.1	698	10	BB327376
10	83	4.1	281	10	BB356081
11	77	3.8	551	17	AZ727531
12	77	3.8	701	17	AZ562461
13	75	3.7	277	10	BB242940
14	73	3.6	312	10	BB215231
15	70	3.4	246	9	AV160656
16	53	2.6	436	14	BQ193951
17	50	2.4	789	10	BE531642
18	47	2.3	335	12	BG209882
19	47	2.3	436	10	BB817265
20	47	2.3	484	12	BG289059
21	47	2.3	636	12	BG773189
22	47	2.3	672	9	AA913512
23	47	2.3	722	13	B1553338
24	47	2.3	813	10	BE387191
25	31	1.5	534	12	BG702763
26	31	1.5	806	12	BE971187
27	31	1.5	936	13	B1560354
28	29	1.4	503	9	A1229932
29	24	1.2	742	13	B1553673
30	23	1.1	384	9	AA373225
31	23	1.1	581	17	AQ976049
32	22	1.1	450	9	A1365378
33	21	1.0	1501	9	AL692789
34	20	1.0	220	17	AZ656375
35	20	1.0	264	9	AV210530
36	20	1.0	296	10	BB551563
37	20	1.0	299	10	BB556696
38	20	1.0	516	17	AZ386497
39	20	1.0	540	17	AQ696374
40	20	1.0	560	17	AZ794922
41	20	1.0	648	10	BB612587
42	20	1.0	795	12	BG481760
43	20	1.0	822	17	BH124617
44	20	1.0	1018	17	CNS07928

ALIGNMENTS

RESULT 1
BB356080
LOCUS
DEFINITION
BB356080 RIKEN full-length enriched, adult male corpus striatum Mus musculus cDNA clone C030011021 3', mRNA sequence.
ACCESSION
BB356080
VERSION
BB356080.2 GI:16405735
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 700)
Arakawa,T., Carinaci,P., Fukuda,S., Futuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Komuro,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jul 12, 2000 this sequence version replaced gi:9067908.

TITLE
JOURNAL
COMMENT
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Tel.: 301 838 0200
Fax: 301 838 0208
Email: szhaettlgr.org
Clones are derived from the mouse BAC library RPT1-23. For BAC library availability, please contact Pieter de Jong

(pieteredejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html
 Plate: 334 row: M column: 15
 Seq primer: 77
 Class: BAC ends.

FEATURES

Source

Location/Qualifiers
 1. 674
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-334M15"
 /clone_11b="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1; EcORI; Site: 2; EcORI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcORI and EcORI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 189 a 174 c 106 g 204 t 1 others
 ORIGIN

Query Match 7.6%; Score 155; DB 17; Length 674;
 Best Local Similarity 100.0%; Pred. No. 8.3e-70;

Matches 155: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1890 GAGCATTCGACATATATCTGGCATATTCGACATCTGCGTACCTGGATCTACAGAT 1949

DB 191 GAGCATTCGACATATATCTGGCATATTCGACATCTGCGTACCTGGATCTACAGAT 132

OY 1950 AAGATATTTGGACACATTCGACATCTGGAACCCGCTGATCTAGCAT 2009

DB 131 AAGATATTTGGACACATTCGACATCTGGAACCCGCTGATCTAGCAT 72

OY 2010 TGAAGAAATTCCTCTCTCTTTGAAGTTGTCAC 2044

DB 71 TGAAGAAATTCCTCTCTCTTTGAAGTTGTCAC 37

RESULT 5
 BH049419/c 834 bp DNA linear GSS 17-JUL-2001
 LOCUS
 DEFINITION
 , DNA sequence.

ACCESSION
 BH049419 GI:14838356

VERSION
 GSS.

KEYWORDS
 house mouse.

SOURCE
 Mus musculus

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 834)

AUTHORS
 Zhou,S., Nierman,W., Malek,J., Shatsman,S., Akintel,B., Levins,M.,
 Tseng,Y.G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.

TITLE
 Mouse BAC End Sequences from Library RPCI-24

JOURNAL
 Unpublished (1999)

COMMENT
 Other_GSSs: RPCI-24-376H14.TV

Contact: Sheng Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end

page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html
 Plate: 376 row: H column: 14
 Seq primer: SP6
 Class: BAC ends.

FEATURES

Source

Location/Qualifiers
 1. 834
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-376H14"
 /clone_11b="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site: 1: BamHI; Site: 2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 249 a 183 c 164 g 238 t
 ORIGIN

Query Match 6.1%; Score 125; DB 17; Length 834;
 Best Local Similarity 100.0%; Pred. No. 5.8e-54;

Matches 125: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 597 CAGCTGTTGGCATCACCCTGGAGCATGACATAGCAACCTTTGTAATAGTGA 656

DB 368 CAGCTGTTGGCATCACCCTGGAGCATGACATAGCAACCTTTGTAATAGTGA 309

OY 657 CAATATGGGCTATGGGCTTACCTTTGACCAAGAAAGCAAGCTATGCGCCGG 716

DB 308 CAATATGGGCTATGGGCTTACCTTTGACCAAGAAAGCAAGCTATGCGCCGG 249

OY 717 ACAGG 721

DB 248 ACAGG 244

RESULT 6
 AV284970 269 bp mRNA linear EST 09-NOV-1999
 LOCUS
 DEFINITION
 AV284970 RIKEN full-length enriched, 11 days pregnant adult female
 ovary and uterus Mus musculus cDNA clone 5031407B10, mRNA sequence.

ACCESSION
 AV284970 GI:6295474

VERSION
 EST.

KEYWORDS
 house mouse.

SOURCE
 Mus musculus

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 269)

AUTHORS
 Konno,H., Aizawa,Y., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
 Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Horii,F.,
 Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai

,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
 Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata

,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Suganara,Y., Suzuki,H.,
 Suzuki,H., Takahashi,F., Tateo,M., Tomioka,N., Tsunoda,Y.,
 Watanabe,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
 Yoshiki,A., Yoshino,M., Yamamatsu,M. and Hayashizaki,Y.

Yoshino,M., Yamamatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Konno,H., et al., 1999)

Unpublished (1999)

TITLE
 JOURNAL
 COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsunumi-Ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp,
 URL: <http://genome-gsc.riken.go.jp/>

Sasaki,N., Izawa,M., Watanabe,M., Tanaka,K., Tanaka,T., Yoneda,Y.,

Db 239 TCCTTTGAAGTTGTCACACTACTAG 263

RESULT 10
BB356081

LOCUS BB356081 281 bp mRNA linear EST 12-JUL-2000

DEFINITION BB356081 RIKEN full-length enriched, adult male corpus striatum Mus musculus cDNA clone C030011022 3', mRNA sequence.

ACCESSION BB356081

VERSION BB356081.1 GI:9067909

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE
AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watanabe, S., Yamamura, T., Yamane, A., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs (Konno, H., et al.)

JOURNAL Unpublished (2000)

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermostabilization and thermoinactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kikuchi, T., Akiyama, J., Shibata, K., Iizawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
source
location/Qualifiers
1..281
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C030011022"
/clone_1lb="RIKEN full-length enriched, adult male corpus striatum"
/sex="male"
/tissue_type="corpus striatum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'

BASE COUNT 81 a 72 c 55 g 73 t

ORIGIN
BamHI

Query Match 4.1%; Score 83; DB 10; Length 281;
Best Local Similarity 100.0%; Pred. No. 5.1e-32;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 608 CCATCCACCCCGTGGAGCATGACATAGCAACCACTTTGGTAATAGTACATATGGGCT 667
|||||
Db 199 CCATCCACCCCGTGGAGCATGACATAGCAACCACTTTGGTAATAGTACATATGGGCT 258
|||||

Qy 668 ATGCGGCTTACCTTTTGAGCAA 690
|||||
Db 259 ATGCGGCTTACCTTTTGAGCAA 281
|||||

RESULT 11
A2727531 551 bp DNA linear GSS 25-JAN-2001

LOCUS A2727531

DEFINITION RPCI-24-127B15.TJ RPCI-24 Mus musculus genomic clone RPCI-24-127B15, DNA sequence.

ACCESSION A2727531

VERSION A2727531.1 GI:12485027

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE
AUTHORS Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 551)
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M., Tesgaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other GSSs: RPCI-24-127B15.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@tigr.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 127 row: B column: 15
Seq primer: SP6
Class: BAC ends.

FEATURES
source
location/Qualifiers
1..551
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-127B15"
/clone_1lb="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pPARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pPARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J DNA.
BASE COUNT 139 a 107 c 117 g 188 t
ORIGIN

Query Match 3.8%; Score 77; DB 17; Length 551;
Best Local Similarity 100.0%; Pred. No. 9.6e-29;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 718 CAGGACCGCTTGTAGCAGCGCTTGTCTATCATCAATCTCGAGACGTGCACCAACATT 777
|||||
Db 82 CAGGACCGCTTGTAGCAGCGCTTGTCTATCATCAATCTCGAGACGTGCACCAACATT 141

Qy 778 CTTGGCCGCGATTGTGT 794
|||||
Db 142 CTTGGCCGCGATTGTGT 158

RESULT 12
A2562461 701 bp DNA linear GSS 20-NOV-2000
LOCUS RPCI-23-207D7.TV RPCI-23 Mus musculus genomic clone RPCI-23-207D7,
DEFINITION DNA sequence.

ACCESSION A2562461 GI:11242281
VERSION A2562461.1
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Zhao, S., Nierman, W., Feildlyum, T., Malek, J., Shatsman, S., Aklnret,
1 (bases 1 to 701)
B., Levis, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C. M.

TITLE Mouse BAC End Sequences from library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-23-207D7.JU
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@edong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (inf@resgen.com). BAC end page:
http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 207 row: D column: 7
Seq primer: 77
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..701
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-207D7"
/clone_1lb="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 220 a 150 c 127 g 204 t
ORIGIN

Query Match 3.8%; Score 77; DB 17; Length 701;

Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1255 TGGCAACGAGATATTTGATGTTGATTTAGCGTGGGCGCTGGCCATAGCTGCT 1314
|||||
Db 701 TGGCAACGAGATATTTGATGTTGATTTAGCGTGGGCGCTGGCCATAGCTGCT 642

Qy 1315 ATCCCTGGGGGAATTAAAC 1331
|||||
Db 641 ATCCCTGGGGGAATTAAAC 625

RESULT 13
BB242940 277 bp mRNA linear EST 06-JUL-2000
LOCUS BB242940
DEFINITION musculus cDNA clone A630094E02 3' similar to D17062 Human HepG2
partial cDNA, clone hmd4c12m5, mRNA sequence.

ACCESSION BB242940 GI:8935686
VERSION BB242940.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Kono, H., Alizawa, K., Akahira, S., Akiyama, J., Arikawa, T., Carninci,
1 (bases 1 to 277)
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, T., Kondo, S., Koya, S., Kurihara, C., Kusabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Oho, T., Owa, C., Saito, H., Sakai, C., Saito, R., Shibata, K., Shibata,
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suganara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,
T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamane, K.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
M., Muramatsu, M. and Hayashizaki, Y.

TITLE Riken Mouse ESTs (Kono, H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gscl.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermolabile and thermostable restriction enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES
source Location/Qualifiers
1..277
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A630094E02"
/clone_1lb="RIKEN full-length enriched, 3 days neonate
thymus"
/tissue-type="thymus"
/dev_stage="3 days neonate"

	Query Match	3.7%	Score 75	DB 10	length 277
	Best Local Similarity	100.0%	Pred. No. 8.4e-26		
	Matches	75	Conservative	0	Mismatches 0
					Indels 0
					Gaps 0
OY	1851	GGACACGAGGTTTATTTGGCACAAGGATCTCGGTTTGACCAATGCACAAATATCTG	1910		
Db	16	GGACACGAGGTTTATTTGGCACAAGGATCTGGTTGACCAATGCACAAATATCTG	75		
OY	1911	GCATATTCACGATAC	1925		
Db	76	GCATATTCACGATAC	90		

```

VERSION      BB215231.1  GI:8880184
KEYWORDS
SOURCE       EST.
             house mouse.

```

TITLE	RIKEN Mouse ESTs (Konno, H., et al.)
JOURNAL	Unpublished (2000)
COMMENT	Contact: Yoshihide Hayashizaki

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenhiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@qsc.riken.go.jp,
url: http://genome.qsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,
'N., Okaaki, Y., Maramatsu, M. and Hayashizaki, Y.
Thermotranscription and thermotranscription of thermolabile enzymes by

FEATURES

SOURCE

BASE COUNT
ORIGIN

QY	480	CCCCCTCCTTTCACCTTCGAAATCCACAGTCAGAGAGAGCAAGTATCTTCAACACAGA	539
Db	104	CCCCCTCCTTTCACCTTCGAAATCCACAGTCAGAGAGAGCAAGTATCTTCAACACAGA	163

RESULT 15
AV160656

KEYWORDS
SOURCE

REFERENCE
AUTHORS

1 (bases 1 to 246)
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Horii, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,

TITLE
JOURNAL
COMMENT
Contact: Chie Owa
Riken Mouse ESTs
Unpublished (1999)
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara,
Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomioka, N.,
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,
Muranatsu, M., Okazaki, Y. and Hayashizaki, Y.

FEATURES	Location/Qualifiers
source	1. .246

BASE COUNT	69 a	53 c	54 g	70 t
ORIGIN				
Query Match	3.4%	Score 70;	DB 9;	Length 246;
Best Local Similarity	100.0%	Pred. No. 3.5e-25;		
Matches 70;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	651	TAGTGACAAT	660			
Db	212	TAGTGACAAT	221			
RESULT 16						
LOCUS	BQ193951/c					
DEFINITION						
ACCESSION	BQ193951					
VERSION	UI-R-CN1-cl2-o-04-0-UI-s1					
KEYWORDS	UI-R-CN1-cl2-o-04-0-UI 3', mRNA sequence.					
SOURCE	BQ193951.1					
ORGANISM	GI:20369502					
REFERENCE	EST.					
AUTHORS	Norway rat.					
TITLE	Rattus norvegicus					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;					
	Rattus.					
	1 (bases 1 to 436)					
	Bonaldo,M.F., Lennon,G. and Soares,M.B.					
	Normalization and subtraction: two approaches to facilitate gene					

discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
COMMENT
Contact: Soares, MB

FEATURES	Location/Qualifiers
source	1. .436

eye library (CV1) corresponding to plates R-CV0-BRH through R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through R-CV1-BTC, and R-CV1-BVO through R-CV1-BVU. This pool represented about 5% of the final driver population. e) A pool of about 10,000 clones from subtracted library B52, BVO and BVOP (7-9.5 kb cDNA library fraction from rat whole embryo), and BXO (0.5-7kb cDNA library fraction from rat whole embryo) corresponding to plates R-B52-BDB through R-B52-BFB, R-BVO-ANK through R-BVO-ANR, R-BVOP-AOI through R-BVOP-AOX, and R-BXO-NOY through R-BXO-ASH. The resulting pool represented 5% of the final driver population. f) a pool of about 7,000 clones from the seven non-normalized libraries that make up the tracer including CYO, CZO, DAO, DBO, DCO, DDO, and DEO corresponding to plates R-CYO-BXP through R-CYO-BXZ, R-CZO-BXA through R-CZO-BYL, R-CZO-BZB-C, R-DAO-BX1 through R-DAO-BXP, R-DAO-BZD through R-DAO-BZH, R-DBO-BYQ through R-DBO-BZA, R-DCO-BZ1 through R-DCO-BZQ, R-DCO-CAY through R-DCO-CBA, R-DDO-BZB through R-DDO-CAA, R-DDO-CBB-C, and R-DEO-CAB through R-DEO-CAL. The resulting pool represented about 10% of the final driver population. g) a pool of about 2,000 clones from the pool of normalized libraries, CNO, that makes up the tracer. The corresponding plates are R-CNO-BKW through R-CNO-BLD, R-CNO-BLG, R-CNO-BLP through R-CNO-BLR, R-CNO-BLT, R-CNO-BLW-X, R-CNO-BMB, and R-CNO-BMF through R-CNO-BML. This pool represented 5% of the final driver population. h) a pool of the 28 most abundant clones in the CNO pool corresponding to the following addresses: bkx-a-09-0-UI, bkx-b-09-0-UI, bkx-b-11-0-UI, bkx-b-10-0-UI, bkx-d-01-0-UI, bkx-d-06-0-UI, bkx-g-08-0-UI, bkx-h-12-0-UI, bkx-a-05-0-UI, bkz-a-06-0-UI, bkz-a-11-0-UI, bkz-c-06-0-UI, bkz-c-09-0-UI, bkz-d-10-0-UI, bla-a-01-0-UI, bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI, bld-a-12-0-UI, bld-f-02-0-UI, bld-c-a-11-0-UI, bld-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, blq-h-04-0-UI, blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5% of the final driver population. i) One abundant CNO clone (corresponding to the address bkz-a-11-0-UI) was digested with Not I and Eco RI and the resulting insert was gel purified. This purified insert was added directly to the driver so that it represented 5% of the final driver population. TAG-lib-UI-R-CN1 TAG-TISSUE-fundus TAG-SEO-TTCGC"

BASE COUNT 107 a 96 c 90 g 143 t
ORIGIN

Query Match 2.6%; Score 53; DB 14; Length 436;
Best Local Similarity 100.0%; Pred. No. 3,9e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1765 AACGAGCCATCAACCTTCTGCTGAGAAATGAGAGACTGTGAGC 1817
|||||
Db 366 AACGAGCCATCAACCTTCTGCTGAGAAATGAGAGACTGTGAGC 314
|||||

RESULT 17
BES31642 789 bp mRNA linear EST 09-AUG-2000
LOCUS 601230651F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3594612 5',
DEFINITION mRNA sequence.
ACCESSION BES31642
VERSION BES31642.1 GI:9760287
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 789)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbstr@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LAM8769 row: e column: 13
High quality sequence stop: 587.
Location/Qualifiers
1..789
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3594612"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 235 a 180 c 225 g 149 t
ORIGIN

Query Match 2.4%; Score 50; DB 10; Length 789;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 852 CATGTGATGGCCAGCGAGCCTGCGACAGACATGTTGAGAGCACACCA 901
|||||
Db 515 CATGTGATGGCCAGCGAGCCTGCGACAGACATGTTGAGAGCACACCA 564
|||||

RESULT 18
B6209882 335 bp mRNA linear EST 21-APR-2001
LOCUS RST29410 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION B6209882
ACCESSION B6209882
VERSION B6209882.1 GI:13731569
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 335)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramchandran,R., Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
21227151
JOURNAL Contact: Scott J. Cain
MEDLINE 3201 Carnegie Ave, Cleveland, OH 44115, USA
COMMENT Athersys, Inc.
Tel: 216 431 9900
Fax: 216 361 9536
Email: scaine@atersys.com
High quality sequence stop: 285.
Location/Qualifiers
1..335
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"

FEATURES
source

/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances."

BASE COUNT 95 a 63 c 74 g 101 t 2 others

ORIGIN

Query Match 2.3%; Score 47; DB 12; Length 335;
Best Local Similarity 100.0%; Pred. No. 5, 1e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1933 CCTGAATCTACAGATTAAGATTTTGGACACATCGAAGCAGGA 1979
|||||
66 CCTGAATCTACAGATTAAGATTTTGGACACATCGAAGCAGGA 112

Db

RESULT 19 436 bp mRNA linear EST 19-NOV-2001
BB817265 RIKEN full-length enriched, lung RCB-0558 LLC CDNA MUS
LOCUS BB817265 musculus cDNA clone G730038N06 3', mRNA sequence.
DEFINITION BB817265.1 GI:16989894
VERSION
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 436)
Akimura,T., Atakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watanahi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE Unpublished (2001)
JOURNAL Contact: Yoshihide Hayashizaki
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanahi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuyama,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES
Location/Qualifiers
1. 436

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G730038N06"
/clone_1lb="RIKEN full-length enriched, lung RCB-0558 LLC CDNA"
/tissue_type="lung"
/cell_line="RCB-0558 LLC"
/note="pooled cell lines: (cell_line=RCB-1751 WEHI 164), (cell_line=RCB-2116 JC), (cell_line=RCB-0035 WEHI-3), (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OH7A), (cell_line=RCB-0559 K-1. F1), (cell_line=RCB-1283 B16 melanoma), (cell_type=B cells, cell_line=RCB-1703 WEHI 231), (cell_type=B cell, cell_line=RCB-2065 MRC-1), (cell_type=Nullipotent stem cell, cell_line=RCB-2070 NE), (tissue_type=bladder, cell_line=RCB-0544 MBT-2), (tissue_type=bone marrow, cell_type=stroma cell, cell_line=RCB-2028 SR-4987), (tissue_type=colon, cell_line=RCB-0549 Cle-H3), (tissue_type=kidney, cell_line=RCB-142 RAG), (tissue_type=submandibular gland, cell_line=RCB-1734 SCA-9 clone 15), (strain=BLB/C, cell_type=B cells, cell_line=RCB-1669 BCL1 clone 13, 20-3B3), (strain=C3H, tissue_type=brain, cell_line=RCB-1443 B3H1)"

BASE COUNT 131 a 116 c 72 g 117 t

ORIGIN

Query Match 2.3%; Score 47; DB 10; Length 436;
Best Local Similarity 100.0%; Pred. No. 5, 6e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 300 CCCAGCAGGGTTTTCATATACACTATATACCTCCGACGAGG 346
|||||
1 CCCAGCAGGGTTTTCATATACACTATATACCTCCGACGAGG 47

Db

RESULT 20 484 bp mRNA linear EST 21-FEB-2001
BG289059 RIKEN full-length enriched, lung RCB-0558 LLC CDNA MUS
LOCUS BG289059 musculus cDNA clone IMAGE:4512680 5', mRNA sequence.
DEFINITION BG289059.1 GI:13044522
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 484)
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-f@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM10398 row: b column: 09
High quality sequence stop: 484.
FEATURES
Location/Qualifiers
1. 484
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4512680"
/clone_1lb="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bladder; Vector: PCMV-SPORE; Site: 1; Note: Site 2: Salt; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.7 kb. Library enriched for

full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."

BASE COUNT 158 a 93 c 107 g 126 t

ORIGIN

Query Match 2.3%; Score 47; DB 12; Length 484;
Best Local Similarity 100.0%; Pred. No. 5.9e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1933 CCTGGAATCTACAGATAAGATATTTGGACACATCGAGACAGCA 1979
|||||
Db 279 CCTGGAATCTACAGATAAGATATTTGGACACATCGAGACAGCA 325
|||||

RESULT 21
Bg773189 636 bp mRNA linear EST 15-MAY-2001
DEFINITION 602721925F1 NIH-MGC_97 Homo sapiens cDNA clone IMAGE:4838671 5',
mRNA sequence.
Bg773189
ACCESSION Bg773189.1 GI:14083842
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 636)
NIH-MGC http://mgi.mcg.mcg.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaab-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshilky and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10773 row: a column: 08
High quality sequence stop: 636.
Location/Qualifiers
1..636
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4838671"
/clone_lib="NIH-MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgaag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
size-selected for average insert size 2.2 kb and
normalized to 5x. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 183 a 131 c 157 g 165 t

ORIGIN

Query Match 2.3%; Score 47; DB 12; Length 636;
Best Local Similarity 100.0%; Pred. No. 6.5e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1933 CCTGGAATCTACAGATAAGATATTTGGACACATCGAGACAGCA 1979
|||||
Db 451 CCTGGAATCTACAGATAAGATATTTGGACACATCGAGACAGCA 497
|||||

RESULT 22
AA913512/C 672 bp mRNA linear EST 10-JUN-1998
LOCUS

DEFINITION 0137h09.s1 Soares_NFL_T-GBC_S1 Homo sapiens cDNA clone
IMAGE:1525697 3' similar to TR:015913 015913 RANDOM SLUG CDNA25
PROTEIN: , mRNA sequence.

ACCESSION AA913512.1 GI:3052904
VERSION
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 672)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
TITLE Unpublished (1997)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaab-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 791 Std Error: 0.00
Seq primer: -40m3 fwd. ET from Amerstham
High quality sequence stop: 470.
Location/Qualifiers
1..672
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1525697"
/clone_lib="Soares_NFL_T-GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCL-CGAP-GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The diver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 176 a 146 c 142 g 208 t

ORIGIN

Query Match 2.3%; Score 47; DB 9; Length 672;
Best Local Similarity 100.0%; Pred. No. 6.6e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1933 CCTGGAATCTACAGATAAGATATTTGGACACATCGAGACAGCA 1979
|||||
Db 203 CCTGGAATCTACAGATAAGATATTTGGACACATCGAGACAGCA 157
|||||

RESULT 23
BI553338 722 bp mRNA linear EST 05-SEP-2001
BI553338
LOCUS 603193475F1 NIH-MGC_95 Homo sapiens cDNA clone IMAGE:5264635 5',
mRNA sequence.
BI553338
ACCESSION BI553338.1 GI:15440650
VERSION
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 722)
NIH-MGC http://mgi.mcg.mcg.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaab-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLAM1167 row: e column: 20
High quality sequence start: 3
High quality sequence stop: 718.
Location/Qualifiers

FEATURES

1. 722

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="5264635"

/clone_lib="NIH_MGC_95"

/tissue_type="hippocampus"

/lab_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified

pBluescript KS+); Site_1: BamHI; Site_2: SalI; XhoI (gtcgag

); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTTNN-3',

size-selected for average insert size 2.5 kb and

normalized to ROT 5. This is a primary library enriched

for full-length clones and constructed using the

Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIH/NHGRI, National

Institutes of Health). Note: this is a NIH_MGC Library."

154 c 168 g 186 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 6.8e-13;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1933 CCTGGAATCTACAGATATGATATTTGGACACATCGAAGCAGCA 1979

|||||

Db

516 CCGGAAATCTACAGATATGATATTTGGACACATCGAAGCAGCA 562

|||||

RESULT 24

BE387191

LOCUS

601277054F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3618180 5',

DEFINITION

mRNA sequence.

ACCESSION

BE387191

VERSION

BE387191.1 GI:9332556

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 813)

AUTHORS

NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC/DCID/DRP

CDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCW286 row: k column: 13

High quality sequence start: 70

High quality sequence stop: 706.

Location/Qualifiers

1. 813

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="3618180"

/clone_lib="NIH_MGC_20"

/tissue_type="melanotic melanoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGCAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 7.1e-13;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1933 CCTGGAATCTACAGATATGATATTTGGACACATCGAAGCAGCA 1979

|||||

Db

409 CCGGAAATCTACAGATATGATATTTGGACACATCGAAGCAGCA 455

|||||

RESULT 25

BG702763

LOCUS

602664640F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4817504 5',

DEFINITION

mRNA sequence.

ACCESSION

BG702763

VERSION

BG702763.1 GI:13974426

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 534)

AUTHORS

NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki

Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10717 row: o column: 09

High quality sequence stop: 534.

Location/Qualifiers

1. 534

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="4817504"

/clone_lib="NIH_MGC_95"

/tissue_type="hippocampus"

/lab_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified

pBluescript KS+); Site_1: BamHI; Site_2: SalI; XhoI (gtcgag

); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTTNN-3',

size-selected for average insert size 2.5 kb and

normalized to ROT 5. This is a primary library enriched

for full-length clones and constructed using the

Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIH/NHGRI, National

Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 0.00016;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1682 TGCAGCCAGCAAACTGGAATACAGAGTGGG 1712

Db 265 TGCAGCCAGCAAACTGATATACAGATGGG 295
 |||
 RESULT 26
 LOCUS BE971187 806 bp mRNA linear EST 04-OCT-2000
 DEFINITION 601651171F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:3934649 5',
 mRNA sequence.
 BE971187
 ACCESSION BE971187.1 GI:10584523
 VERSION EST.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 806)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM7/5 row: m column: 18
 High quality sequence stop: 426.
 Location/Qualifiers
 1..806
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3934649"
 /clone_lib="NIH_MGC_81"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: muscle (skeletal); Vector: pDNR-LTB
 (Clontech); Site_1: SfilI (ggcgagcgccg); Site_2: SfilI
 (ggcgatcattagc); 5' and 3' adaptors were used in cloning
 as follows: 5' adaptor sequence: 5'-CAGCGCATTTATGGCC-3'
 and 3' adaptor sequence:
 5'-ATTCTAGAGCGGCGGCGGCGCATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
 inserts by PCR. This library was enriched for full-length
 clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."

BASE COUNT 222 a 216 c 187 g 181 t
 ORIGIN
 Query Match 1.5%; Score 31; DB 12; Length 806;
 Best Local Similarity 100.0%; Pred. No. 0.00019;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1682 TGCAGCCAGCAAACTGATATACAGATGGG 1712
 |||
 Db 245 TGCAGCCAGCAAACTGATATACAGATGGG 275
 |||
 RESULT 27
 LOCUS B1560354 936 bp mRNA linear EST 05-SEP-2001
 DEFINITION 603253902F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5296080 5',
 mRNA sequence.
 B1560354
 ACCESSION B1560354.1 GI:15447668
 VERSION EST.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 936)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shuraki
 Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM11/49 row: d column: 01
 High quality sequence stop: 666.
 Location/Qualifiers
 1..936
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5296080"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescript (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size selected for average insert size 2.2 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 220 a 252 c 247 g 217 t
 ORIGIN
 Query Match 1.5%; Score 31; DB 13; Length 936;
 Best Local Similarity 100.0%; Pred. No. 0.0002;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1682 TGCAGCCAGCAAACTGATATACAGATGGG 1712
 |||
 Db 524 TGCAGCCAGCAAACTGATATACAGATGGG 554
 |||
 RESULT 28
 LOCUS A1229932 503 bp mRNA linear EST 20-JAN-1999
 DEFINITION EST226627 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
 REMC053 3' end, mRNA sequence.
 A1229932
 ACCESSION A1229932.1 GI:3813819
 VERSION EST.
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 503)
 AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 TITLE Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
 Gene Index
 JOURNAL Unpublished (1998)
 COMMENT Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@igr.org
 Seq primer: M13-21.
 Location/Qualifiers
 1..503
 /organism="Rattus sp."

```

/db_xref="ATCC (Inhost):2037163"
/db_xref="taxon:10118"
/clone="REMCO53"
/clone_1ib="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/notes="vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      152 a      114 c      97 g      140 t
ORIGIN

Query Match      1.4%; Score 29; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1300 TTGGCCATAGCTGCTATCCCTGGGAATT 1328
         |||||||
Db      406 TTGGCCATAGCTGCTATCCCTGGGAATT 378

RESULT 29
BI553673      742 bp      mRNA      linear      EST 05-SEP-2001
LOCUS      603190594F1 NIH_MGC_95 Homo sapiens CDNA clone IMAGE:5262094 5',
DEFINITION      mRNA sequence.
ACCESSION      BI553673
VERSION      BI553673.1 GI:15440985
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE      NIH-MGC http://mgc.ncl.nih.gov/
JOURNAL      1 (bases 1 to 742)
COMMENT      National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs-riemail.nih.gov
              Tissue procurement: Miklos Palkovits, M.D., Ph.D.
              cDNA library preparation: Michael J. Brownstein (NHGRI), Shitaki
              Toshitaki and Piero Carninci (RIKEN)
              DNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLM11660 row: k, column: 23
              High quality sequence stop: 734.
              Location/Qualifiers
                source
                  1..742
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5262094"
                    /clone_1ib="NIH_MGC_95"
                    /tissue_type="hippocampus"
                    /lab_host="DH10B"
                    /note="Organ: brain; Vector: pBluescriptR (modified
                    pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                    ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
                    size-selected for average insert size 2.5 kb and
                    normalized to ROT 5. This is a primary library enriched
                    for full-length clones and constructed using the
                    Cap-trapper method (Carninci, in preparation). Library
                    constructed by M. Brownstein (NIH/NHGRI, National
                    Institutes of Health). Note: this is a NIH-MGC library."
BASE COUNT      211 a      164 c      172 g      195 t
ORIGIN

Query Match      1.2%; Score 24; DB 13; Length 742;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1949 TAAGATATTTGGACACATCGA 1972
         |||||||
Db      554 TAAGATATTTGGACACATCGA 577

```

```

RESULT 30
AA373225      384 bp      mRNA      linear      EST 21-APR-1997
LOCUS      EST852231 HSC172 cells I Homo sapiens CDNA 5' end, mRNA sequence.
DEFINITION      AA373225
ACCESSION      AA373225
VERSION      AA373225.1 GI:2025545
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE      1 (bases 1 to 384)
JOURNAL      Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
MEDLINE      C.J., Lee,N.H., Kitzness,E.F., Weinstein,K.G., Gocayne,J.D., White
COMMENT      O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Mel,C., Clayton,R.A.,
              Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
              ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodex,A.,
              Gnehm,C.L., Hanna,M.C., Heddlom,E., Hinkle,P.S., Jr., Kelley,J.M.,
              Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
              Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
              Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
              Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
              Rederick,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
              Dimke,D., Feng,D.-F., Ferrite,A., Fischer,C., Hastings,G.A., He,M.W.,
              Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
              Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
              Wei,Y.F., Wang,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon
              ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
              Venter,J.C.
              Initial assessment of human gene diversity and expression patterns
              based upon 83 million nucleotides of CDNA sequence
              Nature 377 (6547 Suppl), 3-174 (1995)
              Contact: Kerlavage, AR
              Bioinformatics
              The Institute for Genomic Research
              9712 Medical Center Drive, Rockville, MD 20850 USA
              Tel: 3018699056
              Fax: 3018699423
              Email: arkerlav@tigr.org
              For clone availability, additional sequence and expression
              information related to this EST, please check the TIGR Human Gene
              Index (http://www.tigr.org/tdb/hgi/hgi.html)
              Seq primer: M13 Reverse.
              Location/Qualifiers
                source
                  1..384
                    /organism="Homo sapiens"
                    /db_xref="ATCC (Inhost):177586"
                    /db_xref="taxon:9606"
                    /clone="HSC172 cells I"
                    /cell_type="fibroblast"
                    /cell_line="HSC172 (16PDL)"
                    /dev_stage="fetal"
                    /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
                    ; Site_2: XhoI"
BASE COUNT      101 a      81 c      97 g      101 t      4 others
ORIGIN

Query Match      1.1%; Score 23; DB 9; Length 384;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1780 ACCCTCCTCAGCTGTGAGAAATA 1802
         |||||||
Db      195 ACCCTCCTCAGCTGTGAGAAATA 217

RESULT 31
A0976049      581 bp      DNA      linear      GSS 28-JAN-2000
LOCUS      A0976049/c
DEFINITION      RPCI-23-333M12.TV RPCI-23 Mus musculus genomic clone RPCI-23-333M12

```



```

ACCESSION      , DNA sequence.
                A0976049
VERSION        A0976049.1  GI:6806506
KEYWORDS
SOURCE
ORGANISM       house mouse.
                Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS        Zhao,S., Niemman,W., Feldblum,T., Malek,J., Shatsman,S., Akinret
                ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krul,M., de Jong,P.
                and Fraser,C.M.
TITLE          Mouse BAC End Sequences from Library RPCI-23
JOURNAL
COMMENT        Other_GSSs: RPCI-23-33M12.TJ
                Contact: Shaying Zhao
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: szhao@tigr.org
                Clones are derived from the mouse BAC library RPCI-23. For BAC
                library availability, please contact Pieter de Jong
                (pieter@elgong.med.buffalo.edu). Clones may be purchased from
                BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
                or from Resea ch Genetics (info@resgen.com). BAC end page:
                http://www.tigr.org/tdb/bac_ends/mouse/BAC_end_intro.html
                Plate: 333 row: M column: 12
                Seq primer: T7
                Class: BAC ends.

FEATURES
    source
        Location/Qualifiers
            1..381
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="RPCI-23-33M12"
                /clone_1kb="RPCI-23"
                /sex="Female"
                /lab_host="DH10B"
                /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site:1:
                EcorI; Site:2: EcorI; Female C57BL/6J mouse kidney and/or
                brain genomic DNA was isolated and partially digested
                with a combination of EcorI and EcorI Methylase. Size
                selected DNA was cloned into the pBACe3.6 vector at the
                EcorI sites. The ligation products were transformed into
                DH10B electrocompetent cells (BRL Life Technologies)."
                BASE COUNT      160 a      129 c      130 g      162 t
                ORIGIN
Query Match      1.1%; Score 23; DB 17; Length 581;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1710 GCGAGAGTGGTTGAATTATAT 1732
Db      24 GCGAGAGTGGTTGAAGTATAT 2
RESULT 32
LOCUS      AI365378      450 bp      mRNA      linear      EST 16-FEB-1999
DEFINITION      g008a03.x1 NCI_CGAP-CELL1 Homo sapiens cDNA clone IMAGE:2020876 3'
                similar to h04:J04144 ANGIOTENSIN-CONVERTING ENZYME PRECURSOR,
                SOMATIC (HUMAN);contains MER22.t2 TARI TARI repetitive element ;,
                mRNA sequence.
ACCESSION      AI365378
VERSION        AI365378.1  GI:4125067
KEYWORDS
SOURCE         human.
                Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

REFERENCE	1 (bases 1 to 450)					
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .					
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index					
JOURNAL	Unpublished (1997)					
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: www.bio.llnl.gov/dbdp/image/image.html Insert Length: 601 Std Error: 0.00 Seq primer: -40UP from Gldco High quality sequence stop: 364.					
FEATURES	Location/Qualifiers					
SOURCE	1..450					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="IMAGE:2020876"					
	/clone_id="NCI CGAP_C111"					
	/tissue_type="B-cell, chronic lymphocytic leukemia"					
	/lab_host="DH10B"					
	/note="Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTCAGACGGCAGCGCCGTCATTGTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."					
BASE COUNT	78 a 158 c 150 g 63 t 1 others					
ORIGIN						
Query Match	1.1%; Score 22; DB 9; Length 450;					
Best Local Similarity	100.0%; Pred. No. 8.5;					
Matches	22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	922 CACAAGGCCAAGACTGTATG 943					
DB	317 CAGAAGGCCAAGACTGTATG 338					
RESULT 33						
AL692789/c	1501 bp mRNA linear EST 21-MAR-2002					
LOCUS	AL692789 NAPI Anopheles gambiae cDNA clone NAPI-P03-H-03-5, mRNA sequence.					
DEFINITION	Anopheles gambiae					
ACCESSION	AL692789					
VERSION	AL692789.1 GI:19612674					
KEYWORDS	EST:					
SOURCE	AFrican malaria mosquito.					
ORGANISM	Anopheles gambiae					
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophelinae.					
	1 (bases 1 to 1501)					
REFERENCE	Christophides,G.K., Blass,K., Zdobnov,E.M., Carmouche,R., Benes,V. and Kafatos,F.C.					
AUTHORS	Anopheles gambiae EST, European Molecular Biology Laboratory Unpublished (2002)					
TITLE	Contact: Christophides GK					
JOURNAL	Fotis C. Kafatos Laboratory					
COMMENT	European Molecular Biology Laboratory Meyerhofstrasse 1, 69117 Heidelberg, Germany Tel.: +49 6221 387-440 Fax: +49 6221 387-306 Email: christopheemb@heidelberg.de Plate: P03 row: H column: 03.					

FEATURES
source
Location/Qualifiers
1. .1501
/organism="Anopheles gambiae"
/db_xref="taxon:7165"
/clone="NAP1-P03-H-03-5"
/clone_1lb="NAP1"
/lab_host="E. coli DH10B"
/note="Vector: pT73D-Pac (Pharmacia); Site_1: NotI;
Site_2: EcoRI; ESTs sequenced from the T7 priming site
that reads from the 5' end of cDNA. The NAP1 is a
directionally cloned and normalized, oligo-T primed cDNA
library constructed from a mixture of Anopheles gambiae
developmental stages according to: Bonaldo, Lennon &
Soares (1996): Normalization and Subtraction: Two
Approaches To Facilitate Gene Discovery, Genome Research
6, 791-806."

BASE COUNT 341 a 379 c 352 g 361 t 68 others

ORIGIN

Query Match 1.00; Score 21; DB 9; Length 1501;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 727 TTGTAGACGGCTTGCTTCA 747
DB 399 TTGTAGACGGCTTGCTTCA 379

RESULT 34
AZ656375 220 bp DNA linear GSS 14-DEC-2000
LOCUS 1M0531F19R Mouse 10kb plasmid UUGCIM library Mus musculus genomic
DEFINITION clone UUGCIM0531F19 R, DNA sequence.
ACCESSION AZ656375
VERSION AZ656375.1 GI:11793521
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 220)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0531 Row: F Column: 19
Seq primer: CACACAGAAACAGCATGACAC
Class: Plasmid ends
High quality sequence stop: 220.
Location/Qualifiers
1. .220
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCIM0531F19"
/clone_1lb="Mouse 10kb plasmid UUGCIM library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD22nv. Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (914732114/pLAF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 66 a 46 c 57 g 51 t

ORIGIN

Query Match 1.00; Score 20; DB 17; Length 220;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 937 CTGTATCCCTCTGCTCCCA 956
DB 209 CTGTATCCCTCTGCTCCCA 190

RESULT 35
AV210530 264 bp mRNA linear EST 30-OCT-1999
LOCUS AV210530 RIKEN full-length enriched, adult male testis Mus musculus
DEFINITION cDNA clone 1700124p10 3', mRNA sequence.
ACCESSION AV210530
VERSION AV210530.1 GI:6151067
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 264)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,
Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai,
C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, R., Shibata,
Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,
Suzuki, H., Takahashi, F., Tateo, M., Tomioka, N., Tsunoda, Y.,
Watanabe, S., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,
Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al. 1999)
Unpublished (1999)
Contact: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Sasaki, N., Izawa, M., Watanabe, M., Okazaki, Y., and Hayashizaki,
Y.,
Matsuyama, S., Carninci, P., Muramatsu, M., Okazaki, Y., and Hayashizaki,
Y.,
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7): 3455-3460 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y., and Hayashizaki, Y.,
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5): 463-470 (1999)
Carninci, P., and Hayashizaki, Y.

RESULT 39
A0696374/c 540 bp DNA linear GSS 06-JUL-1999
LOCUS
DEFINITION HS-2171.A2_H01.T7C CIT Approved Human Genomic Sperm Library D Homo
A0696374
ACCESSION sapiens genomic clone Plate-2171 Col=2 Row=0, DNA sequence.
VERSION A0696374.1 GI:5386622
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 540)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2171 row: 0 column: 2
Seq primer: T7
Class: BAC.ends
High quality sequence stop: 540.
Location/Qualifiers
1. 540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate-2171 Col=2 Row=0"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 194 a 130 c 73 g 133 t 10 others
ORIGIN
Query Match 1.0%; Score 20; DB 17; Length 540;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1269 TGTGATGTCAGATTGTTA 1288
Db 417 TGTGATGTCAGATTGTTA 398

RESULT 40
A2794922 560 bp DNA linear GSS 16-FEB-2001
LOCUS
DEFINITION 2M0048119R Mouse 10kb plasmid U0GC1M library Mus musculus genomic
clone U0GC2M0048119 R, DNA sequence.
A2794922
ACCESSION A2794922.1 GI:12941393
VERSION
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 560)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duvul,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL Plasmid Inserts
COMMENT Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0048 row: 1 column: 19
Seq primer: CACACGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 560.
Location/Qualifiers
1. 560
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U0GC2M0048119"
/clone_lib="Mouse 10kb plasmid U0GC1M library"
/sex="male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g14732114[9b]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 151 a 115 c 184 g 110 t
ORIGIN
Query Match 1.0%; Score 20; DB 17; Length 560;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 GGATTGCACGACACTGT 52
Db 321 GGATTGCACGACACTGT 340

RESULT 41
BB612587 648 bp mRNA linear EST 31-AUG-2001
LOCUS
DEFINITION BB612587 RIKEN full-length enriched, 0 day neonate skin Mus
musculus cDNA clone 4632427H21 5', mRNA sequence.
BB612587
ACCESSION BB612587.1 GI:15395169
VERSION
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 648)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,

REFERENCE 1 (bases 1 to 822)
 AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P. and Fraser, C. M.
 TITLE Mouse BAC End Sequences from Library RPCI-24
 JOURNAL Unpublished (1999)
 COMMENT Other_GSSs: RPCI-24-317N21.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhaoc@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.choi.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end page: http://www.tigr.org/db/bac_ends/mouse/bac_end_intro.html
 Plate: 317 row: N column: 21
 Seq primer: SP6
 Class: BAC ends.
 FEATURES
 Location/Qualifiers
 1..822
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-317N21"
 /clone_1lb="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: PTARBAC1; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the PTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."
 BASE COUNT 184 a 182 c 209 g 247 t
 ORIGIN
 Query Match 1.0%; Score 20; DB 17; Length 822;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 502 CCACAGTCAGAGAGCAAG 521
 ||||||||||||||||||
 Db 154 CCACAGTCAGAGAGCAAG 135
 RESULT 44
 CNS07928/c 1018 bp DNA linear GSS 08-JUL-2001
 LOCUS T7 end of clone XBB0AA002C11 of library XBB0AA from strain CBS 4732
 DEFINITION of *Pichia angusta*, genomic survey sequence.
 ACCESSION AL435930
 VERSION AL435930.1 GI:12219343
 KEYWORDS GSS.
 SOURCE *Pichia angusta*.
 ORGANISM *Pichia angusta*.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; *Pichia*.
 REFERENCE 1 (bases 1 to 1018)
 AUTHORS Souciet, J. L., Aigle, M., Artiguenave, F., Blandin, G., Boloitin-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S., de Montigny, J., Dujon, B., Durans, P., Lepingle, A., Llorente, B., Malpertuy, A., Neveuglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wasolowski-Louvel, M., Wincker, P. and Weissenbach, J.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
 MEDLINE 20584771
 PUBMED 11152876
 REFERENCE 2 (bases 1 to 1018)

AUTHORS Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F. and Dujon, B.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia angusta*
 JOURNAL FEBS Lett. 487 (1), 76-81 (2000)
 MEDLINE 20584773
 PUBMED 11152888
 REFERENCE 3 (bases 1 to 1018)
 TITLE Genoscope.
 JOURNAL Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: segref@genoscope.cns.fr - Web: www.genoscope.cns.fr)
 COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
 FEATURES
 Location/Qualifiers
 1..1018
 /organism="Pichia angusta"
 /strain="CBS 4732"
 /db_xref="taxon:4805"
 /clone="XBB0AA002C11"
 /clone_1lb="XBB0AA"
 /note="end : T7"
 /note="complement(<17...>717)"
 /note="similar to Saccharomyces cerevisiae ORF YK1085w (MDH1; malate dehydrogenase precursor, mitochondrial)"
 /note="1 putative frameshift(s)"
 /evidence="not_experimental"
 BASE COUNT 241 a 239 c 315 g 219 t 4 others
 ORIGIN
 Query Match 1.0%; Score 20; DB 17; Length 1018;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1535 TCAGAGACCTTGCTAAGCA 1554
 ||||||||||||||||||
 Db 375 TCAGAGACCTTGCTAAGCA 356
 Search completed: July 5, 2003, 15:48:41
 Job time : 2982 secs

Query Match 100.0%; Score 2049; DB 21; Length 2049;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCAATGCTTACTACATTTGGGAGAGCGGATTTGCACAGACAGTGTACATATC 60
DB 1 TTCAATGCTTACTACATTTGGGAGAGCGGATTTGCACAGACAGTGTACATATC 60
QY 61 AATTTGATGGGCTATGCAAAAATGGCCAGATGCAAGGGGTCTCTCCACAGCGCTTTC 120
DB 61 AATTTGATGGGCTATGCAAAAATGGCCAGATGCAAGGGGTCTCTCCACAGCGCTTTC 120
QY 121 ACCCGTCTTTATCTGTGGCGATCCAGATGGGTCAATGCAATGGCATTTGTGACGTG 180
DB 121 ACCCGTCTTTATCTGTGGCGATCCAGATGGGTCAATGCAATGGCATTTGTGACGTG 180
QY 181 GAATATGATGATTTCCCAAGACAGTGGATGGAGGTCTGAAAGACTAGAGATAA 240
DB 181 GAATATGATGATTTCCCAAGACAGTGGATGGAGGTCTGAAAGACTAGAGATAA 240
QY 241 TATGGCTCTCTGATGAGAGACAATGTTATCTGAGTGCATTCACACAGACTGTGC 300
DB 241 TATGGCTCTCTGATGAGAGACAATGTTATCTGAGTGCATTCACACAGACTGTGC 300
QY 301 CCAGAGAGGTTTTCGAATATACACTCTATATCTGCGCAGAGGAGATTCAGCAACCG 360
DB 301 CCAGAGAGGTTTTCGAATATACACTCTATATCTGCGCAGAGGAGATTCAGCAACCG 360
QY 361 ACCTTTACATACATGCTCTGGATCATGAAGAGCATGATATAGCTACACAAATCTT 420
DB 361 ACCTTTACATACATGCTCTGGATCATGAAGAGCATGATATAGCTACACAAATCTT 420
QY 421 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTGCTAATGTGAGATCAACGAGAC 480
DB 421 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTGCTAATGTGAGATCAACGAGAC 480
QY 481 CCCCTCTCTTACTCTGATTCAGATAGAGAGAGCAAGGATCTTCAAAACAGAC 540
DB 481 CCCCTCTCTTACTCTGATTCAGATAGAGAGAGCAAGGATCTTCAAAACAGAC 540
QY 541 AAGGAATGCTGCTGTAACCTGATGATTTGAATGAGAGACACTTGGTCTTATCAC 600
DB 541 AAGGAATGCTGCTGTAACCTGATGATTTGAATGAGAGACACTTGGTCTTATCAC 600
QY 601 TGGTTTGCATCCACCCTGAGCATGAACATAGACACACTTGTGTAATAGTGCAAT 660
DB 601 TGGTTTGCATCCACCCTGAGCATGAACATAGACACACTTGTGTAATAGTGCAAT 660
QY 661 ATGGGCTATGGGCTTACCTTTTGAAGAAAGAAACAAAGGCTATGCTGGACAG 720
DB 661 ATGGGCTATGGGCTTACCTTTTGAAGAAAGAAACAAAGGCTATGCTGGACAG 720
QY 721 GGACGCTTGTAGACAGGCTTTGCTTCATCAAAATCTCGAGAGCTGTCCCAACATTTT 780
DB 721 GGACGCTTGTAGACAGGCTTTGCTTCATCAAAATCTCGAGAGCTGTCCCAACATTTT 780
QY 781 GGCCTGCTTGTGTCAACACAGAGGAGTCTTGTGACACGACAGACACCTGTCCCAAC 840
DB 781 GGCCTGCTTGTGTCAACACAGAGGAGTCTTGTGACACGACAGACACCTGTCCCAAC 840
QY 841 GGTGGGCTATGATGCAATGGCAGGAGGAGCTGGAACAGACATGTTTGAAGACACAC 900
DB 841 GGTGGGCTATGATGCAATGGCAGGAGGAGCTGGAACAGACATGTTTGAAGACACAC 900
QY 901 ATTATAGACGATCATATCAGAAAGGCAAGAGACTGTATGCTCTGCCAGAG 960
DB 901 ATTATAGACGATCATATCAGAAAGGCAAGAGACTGTATGCTCTGCCAGAG 960
QY 961 GTGACCGGCGCAGTCTTGACAGCTCACCAGTGGGTGAACATGACAGATGTGAGCGTCAG 1020
DB 961 GTGACCGGCGCAGTCTTGACAGCTCACCAGTGGGTGAACATGACAGATGTGAGCGTCAG 1020
QY 1021 CTCAATGCCACACACAGTGAAGAGCTTAACCTGCCCTGAGCTTGTGCGCA 1080

DB 1021 CTCAATGCCACACACAGTGAAGAGCTTAACCTGCCCTGAGCTTGTGCGCA 1080
QY 1081 GGCACAAATGATGAGATTTGCGGCTCAATATTTACACAGGAACTACGGAAGGATCCA 1140
DB 1081 GGCACAAATGATGAGATTTGCGGCTCAATATTTACACAGGAACTACGGAAGGATCCA 1140
QY 1141 TTCTGGACACTCTTCGGACACAGCTCTTGGAAAAACATCTGAAGAGATTTAGAGTGT 1200
DB 1141 TTCTGGACACTCTTCGGACACAGCTCTTGGAAAAACATCTGAAGAGATTTAGAGTGT 1200
QY 1201 CAGAAACCAACCAATCCTGCTCACAGTGAAGAGCTACATACCATCTGGGCA 1260
DB 1201 CAGAAACCAACCAATCCTGCTCACAGTGAAGAGCTACATACCATCTGGGCA 1260
QY 1261 CCAGATATTTGATGATTTGATGATTTGATTTACCGTTGGGCTTGGCCATGCTATCCCT 1320
DB 1261 CCAGATATTTGATGATTTGATGATTTGATTTACCGTTGGGCTTGGCCATGCTATCCCT 1320
QY 1321 GGGGAATTAACACCAATGTCGGAGAGAGATTTGAGGCAATTAAGAAATTTGCA 1380
DB 1321 GGGGAATTAACACCAATGTCGGAGAGAGATTTGAGGCAATTAAGAAATTTGCA 1380
QY 1381 CTTTATGAGATGAAGATATGACCGTGTATTCGAGAGCTTAAGCAATGTTATACAT 1440
DB 1381 CTTTATGAGATGAAGATATGACCGTGTATTCGAGAGCTTAAGCAATGTTATACAT 1440
QY 1441 TACATTTACACATATGAAATATACAGGCTCAGCGGTACGAGAGCATCTCAATCTAT 1500
DB 1441 TACATTTACACATATGAAATATACAGGCTCAGCGGTACGAGAGCATCTCAATCTAT 1500
QY 1501 GGCACACACACCTGCTGCTCATATCCAACTCTTCAGAGACCTTGTAAAGCAATGCT 1560
DB 1501 GGCACACACACCTGCTGCTCATATCCAACTCTTCAGAGACCTTGTAAAGCAATGCT 1560
QY 1561 ACGGACACAGCTGCTGCTCATATCCAACTCTTCAGAGACCTTGTAAAGCAATGCT 1620
DB 1561 ACGGACACAGCTGCTGCTCATATCCAACTCTTCAGAGACCTTGTAAAGCAATGCT 1620
QY 1621 GCTTCACTTATTTCTAATATTTGCGGATAGAGACCAATTTGGCAATTTGGGATGTC 1680
DB 1621 GCTTCACTTATTTCTAATATTTGCGGATAGAGACCAATTTGGCAATTTGGGATGTC 1680
QY 1681 TTGCACGCGCAAAACCTGAATACAGAGTGGGAGAAAGTGTGAAGTTATTTGTAGGC 1740
DB 1681 TTGCACGCGCAAAACCTGAATACAGAGTGGGAGAAAGTGTGAAGTTATTTGTAGGC 1740
QY 1741 GCTAACCCAAAGAAATTCACAGAGAAACAGACCCATCAAACTCTCTACTGTGAGAA 1800
DB 1741 GCTAACCCAAAGAAATTCACAGAGAAACAGACCCATCAAACTCTCTACTGTGAGAA 1800
QY 1801 TACGAGACTCTGTAGCTGACGTGAGATATGTATACGATGCTCTGGAGACGAG 1860
DB 1801 TACGAGACTCTGTAGCTGACGTGAGATATGTATACGATGCTCTGGAGACGAG 1860
QY 1861 TTTTATGCGCAAAAGGATATCTGGGTCTGAGACATGCAACATATATCTGGAATTTCCA 1920
DB 1861 TTTTATGCGCAAAAGGATATCTGGGTCTGAGACATGCAACATATATCTGGAATTTCCA 1920
QY 1921 GATACGCTTACCTGAGATCTACAGAAATPAATATTTTGGACACAATGGAAGAGAA 1980
DB 1921 GATACGCTTACCTGAGATCTACAGAAATPAATATTTTGGACACAATGGAAGAGAA 1980
QY 1981 CTTTGAACCCGCGTGTACTATACATTTTGAAGAAATTTCTTCTCTTTGAAGTGTG 2040
DB 1981 CTTTGAACCCGCGTGTACTATACATTTTGAAGAAATTTCTTCTCTTTGAAGTGTG 2040
QY 2041 ACTACTTAG 2049
DB 2041 ACTACTTAG 2049

RESULT 2

AAA97633
ID AAA97633 standard; DNA; 2271 BP.
XX
AC AAA97633;
XX
DT 02-FEB-2001 (first entry)
XX
DE Mouse neutral/alkaline ceramidase ORF, SEQ ID NO:12.
XX
KM Neutral/alkaline ceramidase; mouse; murine; recombinant production;
KM cellular ceramide content control; antibody; membrane synthesis;
KM lipid engineering; ceramide metabolism; drug development;
KM open reading frame; ORF; ds.
XX
OS Mus sp.
XX
PN MO200058448-A1.
XX
PD 05-OCT-2000.
XX
PF 24-MAR-2000; 2000MO-JP01802.
XX
PR 26-MAR-1999; 99JP-0084743.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Ito M;
XX
DR WPI: 2000-619079/59.
XX P-PSDB: AAB23291.
XX
PT Mammalian neutral/alkaline ceramidase applicable as lipid engineering
PT reagent for studying structure and functions of ceramide as well as
PT developing drugs for diseases associated with ceramide metabolism -
XX
PS Example 4; Page 57-58; 76pp; Japanese.
XX
CC The invention relates to a mouse neutral/alkaline ceramidase and to
CC nucleic acids encoding it. The invention also relates to expression
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
CC ceramidase, the recombinant production of the ceramidase, an antibody
CC against the ceramidase, and a method for controlling the ceramide
CC content in cells and/or tissues by introducing the gene or its antisense
CC nucleic acid into the cells and/or tissues. The ceramidase can be used
CC as a lipid engineering reagent for studying structure and functions of
CC ceramide as well as developing drugs for diseases associated with
CC ceramide metabolism. The present sequence represents the mouse
CC neutral/alkaline ceramidase open reading frame (ORF).
XX
SQ Sequence 2271 BP; 647 A; 541 C; 525 G; 558 T; 0 other;

Query Match 100.0%; Score 2049; DB 21; Length 2271;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCAGTGGCTTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGTACAGATATC 60
DB 223 TTCAGTGGCTTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGTACAGATATC 282
QY 61 AATTGATGGGCTATGGCAAAAATGGCGAATGACAGGGGTCTCTCCACAGGCTGTTC 120
DB 283 AATTGATGGGCTATGGCAAAAATGGCGAATGACAGGGGTCTCTCCACAGGCTGTTC 342
QY 121 AGCGTGGCTTATCTGGCGGATCCAGATGGGTCAAAATCGAATGGCATTTGTGACGCTG 180
DB 343 AGCGTGGCTTATCTGGCGGATCCAGATGGGTCAAAATCGAATGGCATTTGTGACGCTG 402
QY 181 GAACATGATGATTTCCCAACGACGTGAGTGGAGTCTGAAGAGACTAGAGATAA 240
DB 403 GAACATGATGATTTCCCAACGACGTGAGTGGAGTCTGAAGAGACTAGAGATAA 462
QY 241 TATGGCTCTCTATGCAAGACAAATGTTATCTGAGTGCATTTCACACACACTCTGGC 300
DB 463 TATGGCTCTCTATGCAAGACAAATGTTATCTGAGTGCATTTCACACACACTCTGGC 522

DB 463 TATGGCTCTCTATGCAAGACAAATGTTATCTGAGTGCATTTCACACACACTCTGGC 522
QY 301 CCAGCAGGGTTTTTCCAAATATACACTATATCTGCGAGGAGGATTCGCAACGG 360
DB 523 CCAGCAGGGTTTTTCCAAATATACACTATATCTGCGAGGAGGATTCGCAACGG 582
QY 361 ACCTTCACTATAGTCTCTGGGATCATGAGAGACATTGATATAGCTCACAAATCTT 420
DB 583 ACCTTCACTATAGTCTCTGGGATCATGAGAGACATTGATATAGCTCACAAATCTT 642
QY 421 AAACGAGCAAAATCTTTATCAACAAGAAATGTTGCTAATGTGCATCAACGAGC 480
DB 643 AAACGAGCAAAATCTTTATCAACAAGAAATGTTGCTAATGTGCATCAACGAGC 702
QY 481 CCCTCTCTTACCTCTGATCCACAGTCAGAGAGAGCAAGGATTTCTCAACACAGC 540
DB 703 CCCTCTCTTACCTCTGATCCACAGTCAGAGAGAGCAAGGATTTCTCAACACAGC 762
QY 541 AAGAAATGCTGCTTTGAACCTGATGATTTGAATGAGAGACTTGGTCTTATCAC 600
DB 763 AAGAAATGCTGCTTTGAACCTGATGATTTGAATGAGAGACTTGGTCTTATCAC 822
QY 601 TGGTTGCCATCCACCCTGATGACATGAACAATAGCACCTTGTGATATGTCAAAT 660
DB 823 TGGTTGCCATCCACCCTGATGACATGAACAATAGCACCTTGTGATATGTCAAAT 882
QY 661 ATGGGCTATGGGCTTACCTTTTGAACAAGAAACAAAGGCTATGCTGACAG 720
DB 883 ATGGGCTATGGGCTTACCTTTTGAACAAGAAACAAAGGCTATGCTGACAG 942
QY 721 GGACGCTTGTAGCAGAGCTTCTTCATCAAAATCTGAGAGCTGTACCCAAATCTT 780
DB 943 GGACGCTTGTAGCAGAGCTTCTTCATCAAAATCTGAGAGCTGTACCCAAATCTT 1002
QY 781 GGCCCGCATTTGTCTAACACAGGGAGCTTGTGACACGACACAGACCTGTCCAC 840
DB 1003 GGCCCGCATTTGTCTAACACAGGGAGCTTGTGACACGACACAGACCTGTCCAC 1062
QY 841 GGTGGGCTATGATGATGGCGACGGACCTGGCAAGACATGTTTGAAGACACAC 900
DB 1063 GGTGGGCTATGATGATGGCGACGGACCTGGCAAGACATGTTTGAAGACACAC 1122
QY 901 ATTATAGCAGGATCATCTATCAGAAAGGCCAAGAGACTTATGCTCTGCCAGAG 960
DB 1123 ATTATAGCAGGATCATCTATCAGAAAGGCCAAGAGACTTATGCTCTGCCAGAG 1182
QY 961 GTGACCGGCCAGTGTCTACAGTCAACGAGTGGTGAATGACAGATGTGACGCTCAG 1020
DB 1183 GTGACCGGCCAGTGTCTACAGTCAACGAGTGGTGAATGACAGATGTGACGCTCAG 1242
QY 1021 CTCGAATGCCACACACATGTAAGAGCTGTAAACCTGCGCTGGGCTACAGTTTGGCGCA 1080
DB 1243 CTCGAATGCCACACACATGTAAGAGCTGTAAACCTGCGCTGGGCTACAGTTTGGCGCA 1302
QY 1081 GGCACAATTTGATGATTTGGGCTCAATATTTTACACAGGAACTACGGAAGGGATCCA 1140
DB 1303 GGCACAATTTGATGATTTGGGCTCAATATTTTACACAGGAACTACGGAAGGGATCCA 1362
QY 1141 TTCTGGACACCTCTTGGGACACAGCTTTGGGAAACCATCTGAAAGATTTAGAGTGT 1200
DB 1363 TTCTGGACACCTCTTGGGACACAGCTTTGGGAAACCATCTGAAAGATTTAGAGTGT 1422
QY 1201 CAGAAACCCAAACCAATCTGCTTACACATGAGAGCTGACATACCATCTGGGCA 1260
DB 1423 CAGAAACCCAAACCAATCTGCTTACACATGAGAGCTGACATACCATCTGGGCA 1482
QY 1261 CCAGATATTTGATGATTTGATGATTTGATGATTTGATGATTTGATGATTTGATGAT 1320
DB 1483 CCAGATATTTGATGATTTGATGATTTGATGATTTGATGATTTGATGATTTGATGAT 1542
QY 1321 GGGGAATTTAACAACCATGTCGGGACGAGAGATTTCTGAGGCAATTTAAAAAAGATTTGCA 1380
DB 1543 GGGGAATTTAACAACCATGTCGGGACGAGAGATTTCTGAGGCAATTTAAAAAAGATTTGCA 1602

QY 1381 CTTATGGGATGAGATATGACCGTTGTTATCGAGGCTCAAGCAATGTTATACACAT 1440
 Db 1603 CTTATGGGATGAGATATGACCGTTGTTATCGAGGCTCAAGCAATGTTATACACAT 1662
 QY 1441 TACATTACCAATATGAAAGATACCAGGCTCAGGGGTACAGGCGCATCTACAACTAT 1500
 Db 1663 TACATTACCAATATGAAAGATACCAGGCTCAGGGGTACAGGCGCATCTACAACTAT 1722
 QY 1501 GGACACACACCCCTGTCTGCATACATCCAACTCTTCAGAGACCTTGTAAAGCAATTCCT 1560
 Db 1723 GGACACACACCCCTGTCTGCATACATCCAACTCTTCAGAGACCTTGTAAAGCAATTCCT 1782
 QY 1561 ACGGACACAGTAGCCAAACATGAGCAGTGGTCCGAGCCCTCCATCTTCAAAAATCTAATA 1620
 Db 1783 ACGGACACAGTAGCCAAACATGAGCAGTGGTCCGAGCCCTCCATCTTCAAAAATCTAATA 1842
 QY 1621 GCTTCACTTATCTCTAATATTTGCGGATAGACACCAATTTGGCAAAATTTGGGATCTC 1680
 Db 1843 GCTTCACTTATCTCTAATATTTGCGGATAGACACCAATTTGGCAAAATTTGGGATCTC 1902
 QY 1681 TTGACGCGCAGCAAAACCTGAATACAGAGTGGAGAGAGTGTGAAGTTATATTTGTAGC 1740
 Db 1903 TTGACGCGCAGCAAAACCTGAATACAGAGTGGAGAGAGTGTGAAGTTATATTTGTAGC 1962
 QY 1741 GCTAACCCAAAGATTTACGACGAGACCAACCATCAACCTTCTACTGTGAGAA 1800
 Db 1963 GCTAACCCAAAGATTTACGACGAGACCAACCATCAACCTTCTACTGTGAGAA 2022
 QY 1801 TACGAGGACTGTAGCTGAGCTGGCAGATATATATACGATGGCTCTGGGAGACGAG 1860
 Db 2023 TACGAGGACTGTAGCTGAGCTGGCAGATATATATACGATGGCTCTGGGAGACGAG 2082
 QY 1861 TTTTATTTGGCACAAGAAATCTAGGCTCTGAGCAATGCAACAAATATCTGCATATTC 1920
 Db 2083 TTTTATTTGGCACAAGAAATCTAGGCTCTGAGCAATGCAACAAATATCTGCATATTC 2142
 QY 1921 GATACCTGCTACCTGGAATACAGATTAAGATATTTTGGACACATTCGAGACGAGAA 1980
 Db 2143 GATACCTGCTACCTGGAATACAGATTAAGATATTTTGGACACATTCGAGACGAGAA 2202
 QY 1981 CTTCTGAACCCGCTGTCATCTAGCATTTGAGGAATTTCTTCCCTTTTGAAGTTTC 2040
 Db 2203 CTTCTGAACCCGCTGTCATCTAGCATTTGAGGAATTTCTTCCCTTTTGAAGTTTC 2262
 QY 2041 ACTACTTAG 2049
 Db 2263 ACTACTTAG 2271
 RESULT 3
 ID AAA97632 standard; cDNA: 3108 BP.
 AC AAA97632:
 XX 02-FEB-2001 (first entry)
 DE Mouse neutral/alkaline ceramidase clone, SEQ ID NO:11.
 XX Neutral/alkaline ceramidase: mouse; murine; recombinant production;
 KW cellular ceramide content control; antibody; membrane synthesis;
 KW lipid engineering; ceramide metabolism; drug development; ss.
 OS Mus sp.
 PN W0200058448-A1.
 PD 05-OCT-2000.
 PF 24-MAR-2000; 2000WO-JP01802.
 PR 26-MAR-1999; 99JP-0084743.

XX (TAKI), TAKARA SHUZO CO LTD.
 PA Ito M;
 XX WPI: 2000-619079/59.
 DR Mammalian neutral/alkaline ceramidase applicable as lipid engineering
 XX reagent for studying structure and functions of ceramide as well as
 PT developing drugs for diseases associated with ceramide metabolism
 XX Example 4; Page 55-56; 76pp; Japanese.
 PS The invention relates to a mouse neutral/alkaline ceramidase and to
 CC nucleic acids encoding it. The invention also relates to expression
 CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
 CC ceramidase, the recombinant production of the ceramidase, an antibody
 CC against the ceramidase, and a method for controlling the ceramide
 CC content in cells and/or tissues by introducing the gene or its antisense
 CC nucleic acid into the cells and/or tissues. The ceramidase can be used
 CC as a lipid engineering reagent for studying structure and functions of
 CC ceramide as well as developing drugs for diseases associated with
 CC ceramide metabolism. The present sequence represents a clone encoding
 CC mouse neutral/alkaline ceramidase.
 XX Sequence 3108 BP; 888 A; 760 C; 714 G; 746 T; 0 other:
 SQ
 Query Match 100.0%; Score 2049; DB 21; Length 3108;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTCAGTGGCTTACATTTGGCGTTGGAGAGCGGATTTCACAGCAAGTGTACATATTC 60
 Db 948 TTCAGTGGCTTACATTTGGCGTTGGAGAGCGGATTTCACAGCAAGTGTACATATTC 1007
 QY 61 AATTGATGGGCTATGCAAAATGCGCAGATGACGAGGGTCTCCACAGCGCTGTC 120
 Db 1008 AATTGATGGGCTATGCAAAATGCGCAGATGACGAGGGTCTCCACAGCGCTGTC 1067
 QY 121 AGCGTGGCTTATCTTGGCGGATCCAGATGGGTCAAAATGCAATGGCATTTGTGAGCGT 180
 Db 1068 AGCGTGGCTTATCTTGGCGGATCCAGATGGGTCAAAATGCAATGGCATTTGTGAGCGT 1127
 QY 181 GATCATATGATATTTCCCAACGAGTGAAGTTGGAGTCTCGAAGACTAGAGATAA 240
 Db 1128 GATCATATGATATTTCCCAACGAGTGAAGTTGGAGTCTCGAAGACTAGAGATAA 1187
 QY 241 TATGGCTCTGTATCGAAGAGACAAATGTTTCCAGTGGCATTCACACACACTGTGC 300
 Db 1188 TATGGCTCTGTATCGAAGAGACAAATGTTTCCAGTGGCATTCACACACACTGTGC 1247
 QY 301 CCAGCAGGGTTTTCCAATATACACTCTATATCTCGCCAGCGAGGATTCAGCAACCGG 360
 Db 1248 CCAGCAGGGTTTTCCAATATACACTCTATATCTCGCCAGCGAGGATTCAGCAACCGG 1307
 QY 361 ACCTTTCAGTACATATGCTCTGGGATCATGAAGAGACTTGAATATGCTACACAAATCTT 420
 Db 1308 ACCTTTCAGTACATATGCTCTGGGATCATGAAGAGACTTGAATATGCTACACAAATCTT 1367
 QY 421 AACCCAGCAAAATCTTATCAACCAAGGAATGTTGCTAATGTSCAGATCAACCGAGC 480
 Db 1368 AACCCAGCAAAATCTTATCAACCAAGGAATGTTGCTAATGTSCAGATCAACCGAGC 1427
 QY 481 CCTTCCTTACCTTGTGAATTCACAGTCAGAGAGAGAGCAAGATTTCTCAACACAGAC 540
 Db 1428 CCTTCCTTACCTTGTGAATTCACAGTCAGAGAGAGAGCAAGATTTCTCAACACAGAC 1487
 QY 541 AAGGAATGCTGTGTTTAAACTGTGTGATTTGAATGGAAGAGACTTGGGCTTATCAGC 600
 Db 1488 AAGGAATGCTGTGTTTAAACTGTGTGATTTGAATGGAAGAGACTTGGGCTTATCAGC 1547
 QY 601 TGGTTGCATCCACCCGCTGAGTGAACATATGCAACCACTTGTGAATAGTACAT 660

Db 1548 TGGTTGCCATCCACCCGTTGAGCATGAACAATAGCAACCACTTTGTGTAATGTAACAAT 1607
 QY 661 ATGGGCTATGCGGCTTACCTTTTGTGACAAGAAAAGAACAAAGGCTATCGCTGGACAG 720
 Db 1608 ATGGGCTATGCGGCTTACCTTTTGTGACAAGAAAAGAACAAAGGCTATCGCTGGACAG 1667
 QY 721 GGACGGTTTGTAGACAGGCTTTGCTTCATCAAAATCTCGAGAGCTGTCAACCAACATCTT 780
 Db 1668 GGACGGTTTGTAGACAGGCTTTGCTTCATCAAAATCTCGAGAGCTGTCAACCAACATCTT 1727
 QY 781 GGCCGCGATTTGTGTCAACACAGGAGCTTTGTGACACGACACAGCACTGTGCCAAC 840
 Db 1728 GGCCGCGATTTGTGTCAACACAGGAGCTTTGTGACACGACACAGCACTGTGCCAAC 1787
 QY 841 GGCGGCGCTAGCATGTGATGGCGACGCGACCGGCAAGACATGTTTATGAGCAACAC 900
 Db 1788 GGCGGCGCTAGCATGTGATGGCGACGCGACCGGCAAGACATGTTTATGAGCAACAC 1847
 QY 901 ATTATAGACGATCATCTATCAGAAGGCCAAGSAGCTGTATGCCCTTGCTCCAGAG 960
 Db 1848 ATTATAGACGATCATCTATCAGAAGGCCAAGSAGCTGTATGCCCTTGCTCCAGAG 1907
 QY 961 GTGACCGGCCAGCTGCTTGCAGTCAACAGTGGGTGACATGACAGATGTAGCGTCCAG 1020
 Db 1908 GTGACCGGCCAGCTGCTTGCAGTCAACAGTGGGTGACATGACAGATGTAGCGTCCAG 1967
 QY 1021 CTCATGGCCACACACACAGTGAAGAGGTAAACCGCTGGGCTACAGTTTGGCCGA 1080
 Db 1968 CTCATGGCCACACACACAGTGAAGAGGTAAACCGCTGGGCTACAGTTTGGCCGA 2027
 QY 1081 GGCACAATTTGATGAGTTTGGGCGCTCAATATTACACAGGAACTACGAGAGGATCCA 1140
 Db 2028 GGCACAATTTGATGAGTTTGGGCGCTCAATATTACACAGGAACTACGAGAGGATCCA 2087
 QY 1141 TTCTGGGACACTCTTGGGACCAAGCTTTGGGAAAACATCTGAGAGATTTAGAGTCT 1200
 Db 2088 TTCTGGGACACTCTTGGGACCAAGCTTTGGGAAAACATCTGAGAGATTTAGAGTCT 2147
 QY 1201 CAGAAACCCAAACCAATCTGCTCACAGTGGAGAGCTACATACCAATCTTGGCAA 1260
 Db 2148 CAGAAACCCAAACCAATCTGCTCACAGTGGAGAGCTACATACCAATCTTGGCAA 2207
 QY 1261 CCAGATATTTGATGATGTTGAGATTTGATGATGTTGATGATGTTGATGATGTTGATGAT 1320
 Db 2208 CCAGATATTTGATGATGTTGAGATTTGATGATGTTGATGATGTTGATGATGTTGATGAT 2267
 QY 1321 GGGGAATTTACACACATGTCGGGACGAGATTTGTCGAGCAATTTAAAAAAGATTTGCA 1380
 Db 2268 GGGGAATTTACACACATGTCGGGACGAGATTTGTCGAGCAATTTAAAAAAGATTTGCA 2327
 QY 1381 CTTTATGGATGAAGATGATGACCGTGTATTCGACAGCTTAAGCAATTTTATACAT 1440
 Db 2328 CTTTATGGATGAAGATGATGACCGTGTATTCGACAGCTTAAGCAATTTTATACAT 2387
 QY 1441 TACATTACCATATGAAATACAGAGCTCAGCGGTACGAGAGCATCTACAATCTAT 1500
 Db 2388 TACATTACCATATGAAATACAGAGCTCAGCGGTACGAGAGCATCTACAATCTAT 2447
 QY 1501 GGACCAACACACCTGTCTGATACATCAACTCTTTCAGACCTGCTTAAGCAATTTGCT 1560
 Db 2448 GGACCAACACACCTGTCTGATACATCAACTCTTTCAGACCTGCTTAAGCAATTTGCT 2507
 QY 1561 AGGACACAGTGAAGCAACATGAGAGTGGTCCGAGGCCCACTTCTCAAAAATCTAATA 1620
 Db 2508 AGGACACAGTGAAGCAACATGAGAGTGGTCCGAGGCCCACTTCTCAAAAATCTAATA 2567
 QY 1621 GCTTCACTTATCTCAATTTGCGATAGACACCAATTTGGCAAACTTTTGGGATGTC 1680
 Db 2568 GCTTCACTTATCTCAATTTGCGATAGACACCAATTTGGCAAACTTTTGGGATGTC 2627
 QY 1681 TTGACAGCCAGCAAAACCTGAATACAGAGTGGGAGAACTGTTGAGTTATTTTGAAGC 1740
 Db 2628 TTGACAGCCAGCAAAACCTGAATACAGAGTGGGAGAACTGTTGAGTTATTTTGAAGC 2687

QY 1741 GCTAACCAAGATTTACAGAGAGAACACAGACCATCAAACTTCTCACTGTGGAGAA 1800
 Db 2688 GCTAACCAAGATTTACAGAGAGAACACAGACCATCAAACTTCTCACTGTGGAGAA 2747
 QY 1801 TACGAGACTGTGTACCTGACAGATTAATGTATTAAGAGTCTCTGGAGAGAGAG 1860
 Db 2748 TACGAGACTGTGTACCTGACAGATTAATGTATTAAGAGTCTCTGGAGAGAGAG 2807
 QY 1861 TTTTATTTGGCACAAGAGATTTAGGCTGTGAGCAATGCAATATTTAGCATATTTCA 1920
 Db 2808 TTTTATTTGGCACAAGAGATTTAGGCTGTGAGCAATGCAATATTTAGCATATTTCA 2867
 QY 1921 GATACGCTTACCTGTGAATCTACAGATTAATTTTGGACACAAATGGAAGAGAA 1980
 Db 2868 GATACGCTTACCTGTGAATCTACAGATTAATTTTGGACACAAATGGAAGAGAA 2927
 QY 1981 CTCTGAAACCCGCTGCTCACTAGCATTTTGAAGAAATTTCTCTCTTTGAAGTTGTC 2040
 Db 2928 CTCTGAAACCCGCTGCTCACTAGCATTTTGAAGAAATTTCTCTCTTTGAAGTTGTC 2987
 QY 2041 ACTACTTAG 2049
 Db 2988 ACTACTTAG 2996

RESULT 4
 AAA97635
 ID AAA97635 standard; DNA; 4835 BP.
 XX
 AC AAA97635;
 XX
 DT 02-FEB-2001 (first entry)
 DE Mouse neutral/alkaline ceramidase clone, SEQ ID NO:16.
 DE Neutral/alkaline ceramidase; mouse; murine; recombinant production;
 KW cellular ceramide content control; antibody; membrane synthesis;
 KW lipid engineering; ceramide metabolism; drug development; ds.
 XX
 OS Mus sp.
 OS Synthetic.
 OS
 PN MO200058448-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 24-MAR-2000; 2000MO-JP01802.
 XX
 PR 26-MAR-1999; 99JP-0084743.
 XX
 PA (TAKI) TAKARA SHUZO CO LTD.
 XX
 PI Ito M;
 XX
 DR MPI; 2000-619079/59.
 XX
 PT Mammalian neutral/alkaline ceramidase applicable as lipid engineering
 PT reagent for studying structure and functions of ceramide as well as
 PT developing drugs for diseases associated with ceramide metabolism
 XX
 PS Example 6; Page 67-70; 76pp; Japanese.
 CC The invention relates to a mouse neutral/alkaline ceramidase and to
 CC nucleic acids encoding it. The invention also relates to expression
 CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
 CC ceramidase, the recombinant production of the ceramidase, an antibody
 CC against the ceramidase, and a method for controlling the ceramide
 CC content in cells and/or tissues by introducing the gene or its antisense
 CC nucleic acid into the cells and/or tissues. The ceramidase can be used
 CC as a lipid engineering reagent for studying structure and functions of
 CC ceramide as well as developing drugs for diseases associated with
 CC ceramide metabolism. The present sequence represents a clone encoding

CC mouse neutral/alkaline ceramidase.

SQ Sequence 4835 BP; 1460 A; 996 C; 1048 G; 1331 T; 0 other;

Query Match	98.48;	Score 2016;	DB 21;	Length 4835;
100.00	100.00	100.00	100.00	100.00

Best Local Similarity: 100.0%; pred. No. 0;
Matches 2016: Conservative 0; Mismatches

Matches 2016; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY	34	GATTGCACAGACAAAGTGTGAGATATCAATTAATGGGCTATGGCCAAATAATGGCCAGAT	93
Db	696	GATTTCACAGACAAAGTGTGAGATATCAATTAATGGGCTATGGCCAAATAATGGCCAGAT	755
OY	94	GCACGGGGTCTCTCACAGGCGTGTTCAGCCGTCCTTTATCTTGGCGGATCCAGATGGG	153
Db	756	GCACGGGGTCTCTCACAGGCGTGTTCAGCCGTCCTTTATCTTGGCGGATCCAGATGGG	815
OY	154	TCAATATCGAATGGCAATTTGTGAGGCTGGAACCTATGTATGATTTCCCAACACTGAGTTG	213
Db	816	TCAATATCGAATGGCAATTTGTGAGGCTGGAACCTATGTATGATTTCCCAACACTGAGTTG	875
OY	214	GAGGTCCCTGAAGAGACTGAGAGATTAATATGGCTCTCTGATTCGAAGAGCAATGTTTC	273
Db	876	GAGGTCCCTGAAGAGACTGAGAGATTAATATGGCTCTCTGATTCGAAGAGCAATGTTTC	935
OY	274	CTGAGTGCATTCACACACACTCTGGCCACAGAGGTTTTTCCATATATACCTGTATATA	333
Db	936	CTGAGTGCATTCACACACACTCTGGCCACAGAGGTTTTTCCATATATACCTGTATATA	995
OY	334	CTGCCCGACGAGGGAATTCAGCAACCGGACCTTTACGTACATAGTCTGGGATCATGAAG	393
Db	996	CTGCCCGACGAGGGAATTCAGCAACCGGACCTTTACGTACATAGTCTGGGATCATGAAG	1055
OY	394	AGCATTTGATATAGCTCACACAAATCTTTAAACGAGCAAAATCTTTATCAACAAAGGAAT	453
Db	1056	AGCATTTGATATAGCTCACACAAATCTTTAAACGAGCAAAATCTTTATCAACAAAGGAAT	1115
OY	454	GTTCTCTAATGTGCAGATCAACCGAAGCCCTCTCTTACCTTGTGAATCCACAGTCAGAG	513
Db	1116	GTTCTCTAATGTGCAGATCAACCGAAGCCCTCTCTTACCTTGTGAATCCACAGTCAGAG	1175
OY	514	AGACCAAGGTATTTCTTCAAAACAGACAGGAAGAAATGCTGGTCTTGAACCTGGTGATTTG	573
Db	1176	AGACCAAGGTATTTCTTCAAAACAGACAGGAAGAAATGCTGGTCTTGAACCTGGTGATTTG	1235
OY	574	AATGGAGAGAAGCTTGGGTCTTATAGCTGTGTTGCCATCCACCCCGTAGCATGAACAT	633
Db	1236	AATGGAGAGAAGCTTGGGTCTTATAGCTGTGTTGCCATCCACCCCGTAGCATGAACAT	1295
OY	634	AGCAACCACTTTTGTAATATGTGACAAATATGGGCTATGCGGCTTACCTTTTGTAGCAAGAA	693
Db	1296	AGCAACCACTTTTGTAATATGTGACAAATATGGGCTATGCGGCTTACCTTTTGTAGCAAGAA	1355
OY	694	AAGAAACAAAGGCTATGCGGCTGAGACAGGAGACGTTTGTAGAGAGCTTGTGCATCAAT	753
Db	1356	AAGAAACAAAGGCTATGCGGCTGAGACAGGAGACGTTTGTAGAGAGCTTGTGTCAATCAAT	1415
OY	754	CTCGGAGACGTGTACCCCAACATCTTGGCCGCAATTTGTCAACACAGGGAGACTGTGT	813
Db	1416	CTCGGAGACGTGTACCCCAACATCTTGGCCGCAATTTGTCAACACAGGGAGACTGTGT	1475
OY	814	GACAAACGACAAGACACCTGTGCCAACGCGTGGGCTTAGCATGTGCATGGCCAGCGGACT	873
Db	1476	GACAAACGACAAGACACCTGTGCCAACGCGTGGGCTTAGCATGTGCATGGCCAGCGGACT	1535
OY	874	GGACAAAGCATGTTTGAAGAGCACACATTTATAGACGGAATCATCTATCAAAAGGCCAAG	933
Db	1536	GGACAAAGCATGTTTGAAGAGCACACATTTATAGACGGAATCATCTATCAAAAGGCCAAG	1595
OY	934	GAGCTGTATGCTCTGACCTCCACAGAGATGACCGGCCAGTAGTGTGACAGCTCACACAGG	993
Db	1596	GAGCTGTATGCTCTGACCTCCACAGAGATGACCGGCCAGTAGTGTGACAGCTCACACAGG	1655
OY	994	GTGAACATGACAGATGTGAGCGTCCAGCTCAATGCCACACACAGAGTGAAGAGCTGTAAA	1053

Db	1656	GTGAACATGACAGATGTGAGCGCTCAGCTCAATGCCACACACAGTGAAGACGTGTAA	1715
Qy	1054	CCTGCGCCCTGAGCTACAGTTTTCGCCGACGACCAATGTATGAGATTTGCGGCTCAATATT	1113
Db	1716	CCCTGCCCTGGGCTACAGTTTTCGCCGACGACCAATGTATGAGATTTGCGGCTCAATATT	1775
Qy	1114	ACACAGGAACCTACGGAAGGGGATCCATTCTGGAGACCTCTTGGGACACAGCTCTGGGA	1173
Db	1776	ACACAGGAACCTACGGAAGGGGATCCATTCTGGAGACCTCTTGGGACACAGCTCTGGGA	1835
Qy	1174	AAACCATCTGAAGAGATTGTAGAGTGTCAAGAAACCAACCAATCTGCTTCACAGTGA	1233
Db	1836	AAACCATCTGAAGAGATTGTAGAGTGTCAAGAAACCAACCAATCTGCTTCACAGTGA	1895
Qy	1234	GAGCTGACGATACCAATCCTTGGCACCCAGATTTGTGATGTTCAGATTGTACCGGTT	1293
Db	1896	GAGCTGACGATACCAATCCTTGGCACCCAGATTTGTGATGTTCAGATTGTACCGGTT	1955
Qy	1294	GGGTCTCTTGGCCATAGCTGCTATTCCCTGGGGAATTAACAACATGTCGGGACGAAGATT	1353
Db	1956	GGGTCTCTTGGCCATAGCTGCTATTCCCTGGGGAATTAACAACATGTCGGGACGAAGATT	2015
Qy	1354	CCTGACGCAATTAAGAAATTTTGCACTTATGCGATGAAGATATAGCCGTGTATC	1413
Db	2016	CCTGACGCAATTAAGAAATTTTGCACTTATGCGATGAAGATATAGCCGTGTATC	2075
Qy	1414	GCAGGTCTAACCAATGTTTATACACATTTACATTCACCATATGTAAGATAACAGGCTCAG	1473
Db	2076	GCAGGTCTAACCAATGTTTATACACATTTACATTTACACCATATGTAAGATAACAGGCTCAG	2135
Qy	1474	CGGTACGACGACGATCTACAAATCTATGAGACCAACACACCCTGTGTGCATACATCCACTC	1533
Db	2136	CGGTACGACGACGATCTACAAATCTATGAGACCAACACACCCTGTGTGCATACATCCACTC	2195
Qy	1534	TTTCAGAGACCTTGCTGAAGGCAATTGCTACGAGACAGTAGGCAACATGAGCAGTGGTCC	1593
Db	2196	TTTCAGAGACCTTGCTGAAGGCAATTGCTACGAGACAGTAGGCAACATGAGCAGTGGTCC	2255
Qy	1594	GAGCCTCATTTCTTCAAAAATCTAATAGCTTCACTTATTCCTAATATTGGGATATAGCA	1653
Db	2256	GAGCCTCATTTCTTCAAAAATCTAATAGCTTCACTTATTCCTAATATTGGGATATAGCA	2315
Qy	1654	CCAATTGGCAAAATTTTGGGGATGTCTTGAGCAGACGAAACCTGAAATACAGAGTGGGA	1713
Db	2316	CCAATTGGCAAAATTTTGGGGATGTCTTGAGCAGACGAAACCTGAAATACAGAGTGGGA	2375
Qy	1714	GAAGTGTTGAAGTTATATTTGTAGGGGCTTAACCAAGAATTCAGCAGACACGACCC	1773
Db	2376	GAAGTGTTGAAGTTATATTTGTAGGGGCTTAACCAAGAATTCAGCAGACACGACCC	2435
Qy	1774	CATCAAAACCTTCTCAGCTGTGGAGAAATACAGGACCTCTGAGCTGAGCTGGCAGATTAATG	1833
Db	2436	CATCAAAACCTTCTCAGCTGTGGAGAAATACAGGACCTCTGAGCTGAGCTGGCAGATTAATG	2495
Qy	1834	TATAACGATGCTCTCTGGAGACGAGGTTTTATTGGCACAAAAGAAATACGTGGTCTGAGC	1893
Db	2496	TATAACGATGCTCTCTGGAGACGAGGTTTTATTGGCACAAAAGAAATACGTGGTCTGAGC	2555
Qy	1894	AATGCAACAATATCTGGCAATTCGAGATCTGCTACCCCTGGAATCTACAGAAATAGA	1953
Db	2556	AATGCAACAATATCTGGCAATTCGAGATCTGCTACCCCTGGAATCTACAGAAATAGA	2615
Qy	1954	TATTTTGGACAACAATCGGACAGAGAACTTCTGAACCCCGCTGTCAATACTGCAATTTGAA	2013
Db	2616	TATTTTGGACAACAATCGGACAGAGAACTTCTGAACCCCGCTGTCAATACTGCAATTTGAA	2675
Qy	2014	GGAATTTCTTCTCTTTTGAAGTGTCTACTACTTAG 2049	
Db	2676	GGAATTTCTTCTCTTTTGAAGTGTCTACTACTTAG 2711	

RESULT 5

AAH98033
ID AAH98033 standard; DNA: 551 BP.
XX
AC AAH98033:
XX
DT 10-OCT-2001 (first entry)
XX
DE Murine 7-transmembrane G-protein coupled receptor coding sequence #277.
XX
KM Murine: stromal stem cell; signalling; vaccine; 7TM-GPCR;
7-transmembrane G-protein coupled protein receptor; ds.
XX
OS Mus sp.
XX
PN MO200160999-A1.
XX
PD 23-AUG-2001.
XX
PF 14-FEB-2001; 2001MO-0504700.
XX
PR 14-FEB-2000; 2000US-0182377.
XX
PA (IMCL-) IMCLONE SYSTEMS INC.
(UYPR-) UNIV PRINCETON.
XX
PI Lemischka IR, Witte L, Pereira DS;
XX
DR WPI; 2001-522596/57.
XX
PT DNA Sequences encoding 7-transmembrane G-protein coupled protein
PT receptors characteristic of hematopoietic stem cells, useful for
PT treating leukemia -
XX
PS Claim 1; Page 125; 176pp; English.
XX
CC The present invention relates to murine coding sequences for
CC 7-transmembrane G-protein coupled protein receptors (7TM-GPCRs). The
CC present sequence is one such murine 7TM-GPCR coding sequence. The present
CC sequence was derived from stromal stem cells. The present sequence
CC and its corresponding protein are useful in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate 7TM-GPCR expression.
CC 7TM-GPCRs identify specific signalling molecules, to activate an
CC effector-signalling cascade that triggers an intracellular response and
CC eventually a biological effect.
XX
SQ Sequence 551 BP; 145 A; 135 C; 150 G; 120 T; 1 other:
Query Match 16.1%; Score 329; DB 22; Length 551;
Best Local Similarity 99.8%; Pred. No. 6.1e-156;
Matches 449; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 802 GGGGACTCTTGTACACAGCAGACAGACACTGTCCTCCAGAGGTGGGCTAGCATGTGCATG 861
DB 1 GGGGACTCTTGTGACACAGCAGACAGACACTGTCCTCCAGAGGTGGGCTAGCATGTGCATG 60
QY 862 GCCAGCGACCTGACAGACATGTTGAGAGACACACATTTAGAGCGATCATCTAT 921
DB 61 GCCAGCGACCTGACAGACATGTTGAGAGACACACATTTAGAGCGATCATCTAT 120
QY 922 CAGAGGCCAAGAGAGCTGTATGCTCTGCTCCAGAGAGTGACCGCCAGTGTTCGA 981
DB 121 CAGAGGCCAAGAGAGCTGTATGCTCTGCTCCAGAGAGTGACCGCCAGTGTTCGA 180
QY 982 GCTCACCAGTGGGTGAAATGACAGATGAGAGGCTCCAGTCAATGCCACACACAGTG 1041
DB 181 GCTCACCAGTGGGTGAAATGACAGATGAGAGGCTCCAGTCAATGCCACACACAGTG 240
QY 1042 AAGACGTGTAAACCTCCCTGGCTACAGTTTGGCGAGCACAATTGATGAGTTTG 1101
DB 241 AAGACGTGTAAACCTCCCTGGCTACAGTTTGGCGAGCACAATTGATGAGTTTG 299
QY 1102 GGCCTCAATTTACACAGGAGACTACGAAAGGAGATTCATTTGGGACACTCTTCGGGAC 1161
|||||

DB 300 GGCCTCAATTTACACAGGAGACTACGAAAGGAGATTCATTTGGGACACTCTTCGGGAC 359
QY 1162 CAGCTCTTGGGAAACCATCTGAAAGATTTGATGAGTGTGAGAAACCAACCAATCTG 1221
DB 360 CAGCTCTTGGGAAACCATCTGAAAGATTTGATGAGTGTGAGAAACCAACCAATCTG 419
QY 1222 CTTACAGTGGAGAGCTGACATACACAT 1251
DB 420 CTTACAGTGGAGAGCTGACATACACAT 449
RESULT 6
AAF9977
ID AAF9977 standard; DNA: 2283 BP.
XX
AC AAF9977;
XX
DT 20-JUL-2001 (first entry)
XX
DE DNA encoding rat protein involved in acid amide bond hydrolysis.
XX
KM Rat; acid amide bond hydrolysis; ceramide; cytosolic; cancer; ds.
XX
OS Rattus norvegicus.
XX
FH Key location/Qualifiers
FT CDS 1..2283
FT /tag- a
FT /partial
FT /note="this sequence does not contain a stop codon"
PN JP2001057890-A.
XX
PD 06-MAR-2001.
XX
XX 23-AUG-1999; 99JP-0235218.
PF 23-AUG-1999; 99JP-0235218.
PR (MITU) MITSUBISHI CHEM CORP.
XX
PA WPI; 2001-304133/32.
XX
DR P-PSDB; AAB97029.
DR
XX
PT Novel protein, used to treat abnormal cell growth e.g. in cancers, is
PT capable of hydrolyzing the acid amide bond between sphingosine and the
PT fatty acid in ceramide -
XX
PS Claim 3; Page 9-11; 14pp; Japanese.
XX
CC The present sequence is provided in a specification relating to a
CC protein having a 761 residue amino acid sequence or its variant
CC comprising a replacement, deletion, insertion, addition or reversion of
CC at least one amino acid, but retaining the activity of hydrolyzing the
CC acid amide bond between sphingosine and the fatty acid in a ceramide.
CC The protein can be used to treat diseases caused by abnormal cell
CC growth such as cancers.
XX
SQ Sequence 2283 BP; 641 A; 584 C; 524 G; 534 T; 0 other:
Query Match 3.3%; Score 68; DB 22; Length 2283;
Best Local Similarity 100.0%; Pred. No. 8.3e-24;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1057 GCCCTGGGCTACAGTTTGGCGAGCAGACATTTGATGAGTTTGGGCTCAATATTACA 1116
DB 1294 GCCCTGGGCTACAGTTTGGCGAGCAGACATTTGATGAGTTTGGGCTCAATATTACA 1353
QY 1117 CAGGGAAC 1124
DB 1354 CAGGGAAC 1361
|||||

RESULT 7
ABN53673
ID ABN53673 standard; DNA; 65 BP.
XX
XX
AC ABN53673;
XX
XX
DT 15-JUL-2002 (first entry)
XX
XX
DE Mouse spliced transcript detection oligonucleotide SEQ ID NO:26421.
XX
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
XX
OS Mus musculus.
XX
XX
PN WO200210449-A2.
XX
XX
PD 07-FEB-2002.
XX
XX
PF 20-JUL-2001; 2001WO-1B01903.
XX
XX
PR 28-JUL-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
XX
PA (COMP-) COMPUGEN INC.
XX
XX
PI Shoshan A, Masserman A, Mintz E, Mintz L, Faigler S;
XX
XX
DR WPI; 2002-257383/30.
XX
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes
XX
XX
PS Example 1; SEQ ID 26421; 47pp; English.
XX
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridizing selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59389 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_Pct_sequences.
XX
XX
SQ Sequence 65 BP; 20 A; 18 C; 14 G; 13 T; 0 other;

Query Match 3.2%; Score 65; DB 24; Length 65;
Best Local Similarity 100.0%; Pred No. 2.8e-22;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1767 CCAGACCATCAAAACCTTCTCCTGAGAGAAATACGAGACTGTAGCTGACTGGCA 1826
DB 1 CCAGACCATCAAAACCTTCTCCTGAGAGAAATACGAGACTGTAGCTGACTGGCA 60
OY 1827 GATTA 1831

DB 61 GATTA 65
|||||

RESULT 8
AAF99979
ID AAF99979 standard; CDNA; 539 BP.
XX
XX
AC AAF99979;
XX
XX
DT 20-JUL-2001 (first entry)
XX
XX
DE PCR product isolated from rat CDNA.
XX
XX
KW Rat; acid amide bond hydrolysis; ceramide; cytosolic; cancer; ss.
KW Rattus norvegicus.
XX
XX
PN JP2001057890-A.
XX
XX
PD 06-MAR-2001.
XX
XX
PF 23-AUG-1999; 99JP-0235218.
XX
XX
PR 23-AUG-1999; 99JP-0235218.
XX
XX
PA (MITU) MITSUBISHI CHEM CORP.
XX
XX
DR WPI; 2001-304133/32.
XX
XX
PT Novel protein, used to treat abnormal cell growth e.g. in cancers, is
PT capable of hydrolyzing the acid amide bond between sphingosine and the
PT fatty acid in ceramide
XX
XX
PS Example 3; Page 14; 14pp; Japanese.
XX
XX
CC The present sequence is provided in a specification relating to a
CC protein having a 761 residue amino acid sequence or its variant
CC comprising a replacement, deletion, insertion, addition or reversion of
CC at least one amino acid, but retaining the activity of hydrolysing the
CC acid amide bond between sphingosine and the fatty acid in a ceramide.
CC The protein can be used to treat diseases caused by abnormal cell
CC growth such as cancers.
XX
XX
SQ Sequence 539 BP; 181 A; 110 C; 113 G; 135 T; 0 other;

Query Match 2.6%; Score 53; DB 22; Length 539;
Best Local Similarity 100.0%; Pred No. 3.3e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1765 AACGAGACCATCAAAACCTTCTCCTGAGAGAAATACGAGACTGTAGC 1817
DB 154 AACGAGACCATCAAAACCTTCTCCTGAGAGAAATACGAGACTGTAGC 206

RESULT 9
AAD22668
ID AAD22668 standard; CDNA; 2507 BP.
XX
XX
AC AAD22668;
XX
XX
DT 26-FEB-2002 (first entry)
XX
XX
DE Human ceramidase CDNA.
XX
XX
KW Human; ceramidase; nephrotropic; antipsoriatic; antitense therapy;
KW gene therapy; proliferative disorder; cancer; cardiovascular disease;
KW inflammation; neurodegenerative disorder; cytostatic; immunosuppressive;
KW neurotropic; signal transduction; breast cancer; autoimmune disorder;
KW Alzheimer's disease; growth deficiency; lesion; lupus nephritis;
KW glomerular disease; ss.
XX
XX
OS Homo sapiens.


```
XX Key Location/Qualifiers
FH 39..2324
FT /tag= a
FT /product= "Human ceramidase protein"
FT /note= "There is an additional stop codon from position
FT 2325-2327"
FT 39..95
FT /tag= b
FT mat_peptide 96..2321
FT /tag= c
FT /product= "Mature ceramidase protein"
XX
XX WO200155410-A2.
XX
XX 02-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US02866.
XX
XX 28-JAN-2000; 2000US-178975P.
XX
XX (MUSC-) MUSC FOUND RES DEV.
XX
XX Hannun, YA, El Bawab S;
XX
XX WPI: 2002-025687/03.
XX
XX P-PSDB: AAE13544.
XX
XX Human mitochondrial ceramidase protein and gene, modulation of which is
XX useful for preventing and treating proliferative disorders e.g. cancer,
XX cardiovascular disease, inflammation and neurodegenerative disorders -
XX
XX Claim 1; Fig 3; 113pp; English.
XX
XX The present sequence is a cDNA encoding human mitochondrial ceramidase
XX protein. Ceramidase protein and gene are useful for treating a disease
XX or disorder involving cell over proliferation or sphingolipid signal
XX transduction especially breast cancer, cardiovascular disorder or
XX inflammation. Ceramidase protein and gene are useful for treating
XX disorders involving deficient cell proliferation or growth e.g.
XX neurodegenerative disorders (Alzheimer's disease), growth deficiencies
XX and lesions. Ceramidase protein is also useful for diagnosis of
XX hyperproliferative diseases. Ceramidase gene can be used as an immunogen
XX to generate antibodies which are useful for diagnosis, prevention and
XX treatment of hyperproliferative diseases and for detecting ceramidase
XX gene product in a biological sample. The hyperproliferative disorders
XX include cancers and autoimmune disorders such as lupus nephritis,
XX glomerular disease. Ceramidase gene is also useful in antisense therapy.
XX
XX Sequence 2507 BP; 720 A; 583 C; 567 G; 637 T; 0 other:
XX
XX Query Match 2.3%; Score 47; DB 24; Length 2507;
XX Best Local Similarity 100.0%; Pred. No. 3.5e-13;
XX Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1933 CCTGAATCTACAGATAAGATATTTGGACACAAATCGAGACGAGA 1979
Db 2208 CCTGAATCTACAGATAAGATATTTGGACACAAATCGAGACGAGA 2254
XX
XX RESULT 10
XX ABL99947
XX ID ABL99947 standard; CDNA: 3058 BP.
XX
XX ABL99947;
XX
XX 21-AUG-2002 (first entry)
XX
XX Human stialyl transferase (HST30-1) 11.99 encoding cDNA SEQ ID NO 1.
XX
XX Human; stialyl transferase; HST30-1; enzyme; cytosolic; virucidal;
XX immunomodulatory; antiinflammatory; haemostatic; malignant tumour;
```

```
KW human immunodeficiency virus; HIV; infection; immunological disease;
KW gene therapy; gene; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 1312..1641
FT CDS /tag= a
FT /product= "stialyl transferase (HST30-1) 11.99"
FT /note= "Claimed in claim 6"
XX
XX WO200233075-A1.
XX
XX 25-APR-2002.
XX
XX 25-JUN-2001; 2001WO-CN01058.
XX
XX 28-JUN-2000; 2000CN-0116809.
XX
XX (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX
XX Mao Y, Xie Y;
XX
XX WPI: 2002-340235/37.
XX
XX P-PSDB: ABB77549.
XX
XX Human stialyl transferase (HST30-1) 11.99 and encoding polynucleotide,
XX used in diagnosis and treatment of malignant tumors, hemopathy, human
XX immunodeficiency virus infection, immunological diseases and
XX inflammation -
XX
XX Claim 6; Page 31-33; 39pp; Chinese.
XX
XX The invention relates to human stialyl transferase (HST30-1) 11.99 with
XX cytosolic, virucidal, immunomodulatory, antiinflammatory and
XX haemostatic activity. The protein and encoding polynucleotide are used
XX in diagnosis and treatment of malignant tumour, hemopathy, human
XX immunodeficiency virus (HIV) infection, immunological diseases and
XX various inflammations. The polynucleotide is useful in gene therapy.
XX
XX Sequence 3058 BP; 932 A; 544 C; 597 G; 985 T; 0 other:
XX
XX Query Match 2.1%; Score 43; DB 24; Length 3058;
XX Best Local Similarity 100.0%; Pred. No. 3.8e-11;
XX Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1937 GAATCTACAGATAAGATATTTGGACACAAATCGAGACGAGA 1979
Db 519 GAATCTACAGATAAGATATTTGGACACAAATCGAGACGAGA 561
XX
XX RESULT 11
XX AAA97627
XX ID AAA97627 standard; DNA: 38 BP.
XX
XX AAA97627;
XX
XX 02-FEB-2001 (first entry)
XX
XX Mouse neutral/alkaline ceramidase PCR primer, SEQ ID NO:6.
XX
XX Neutral/alkaline ceramidase; mouse; murine; recombinant production;
XX cellular ceramide content control; antibody; membrane synthesis;
XX lipid engineering; ceramide metabolism; drug development; PCR primer; ss.
XX
XX Mus sp.
XX
XX WO200058448-A1.
XX
XX 05-OCT-2000.
XX
XX 24-MAR-2000; 2000WO-JP01802.
XX
```

PR 26-MAR-1999; 99JP-0084743.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX Ito M;
XX
XX WPI; 2000-619079/59.
XX
XX Mammalian neutral/alkaline ceramidase applicable as lipid engineering
PT reagent for studying structure and functions of ceramide as well as
PT developing drugs for diseases associated with ceramide metabolism
XX
XX Example 3; Page 53; 76pp; Japanese.
XX
XX The invention relates to a mouse neutral/alkaline ceramidase and to
CC nucleic acids encoding it. The invention also relates to expression
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
CC ceramidase, the recombinant production of the ceramidase, an antibody
CC against the ceramidase, and a method for controlling the ceramide
CC content in cells and/or tissues by introducing the gene or its antisense
CC nucleic acid into the cells and/or tissues. The ceramidase can be used
CC as a lipid engineering reagent for studying structure and functions of
CC ceramide as well as developing drugs for diseases associated with
CC ceramide metabolism. The present sequence represents a PCR primer
CC used to isolate nucleic acids encoding mouse neutral/alkaline
CC ceramidase.
XX
XX Sequence 38 BP; 8 A; 11 C; 9 G; 10 T; 0 other:
SQ
Query Match 1.9%; Score 38; DB 21; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 733 GCAGGCTTGCTTCATCAATCTCGAGACGTGCACC 770
DB 1 GCAGGCTTGCTTCATCAATCTCGAGACGTGCACC 38
RESULT 12
AAC03574
ID AAC03574 standard; CDNA; 299 BP.
XX
XX AAC03574;
AC
XX 06-OCT-2000 (first entry)
DT
XX
XX Human secreted protein 5' EST, SEQ ID NO: 3572.
DE
XX
XX Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX EP1033401-A2.
PN
XX 06-SEP-2000.
PD
XX 21-FEB-2000; 2000EP-0200610.
PE
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GEST) GENSET.
PA
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX WPI; 2000-500381/45.
DR
XX P-PDB; AAC03568.
DR
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 3572; 71pp + CD-ROM; English.

XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
XX Sequence 299 BP; 84 A; 68 C; 76 G; 68 T; 3 other:
SQ
Query Match 1.5%; Score 31; DB 21; Length 299;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1682 TGCACGCCAGCAAACTGATACAGAGTGGC 1712
DB 264 TGCACGCCAGCAAACTGATACAGAGTGGC 294
RESULT 13
AAZ06239
ID AAZ06239 standard; DNA; 1194 BP.
XX
XX AAZ06239;
AC
XX 30-SEP-1999 (first entry)
DT
XX
XX Human secreted protein gene No. 21.
DE
XX
XX Human: secreted protein; fusion protein; gene therapy; protein therapy;
KM diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KM developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KM immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KM inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
XX Homo sapiens.
OS
XX WO9935158-A1.
PN
XX 15-JUL-1999.
PD
XX 06-JAN-1999; 99WO-US00108.
PE
XX 07-JAN-1998; 98US-0070704.
PR 07-JAN-1998; 98US-0070657.
PR 07-JAN-1998; 98US-0070658.
PR 07-JAN-1998; 98US-0070692.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
XX
XX WPI; 1999-444190/37.
DR P-PDB; AAY38406, AAY38464, AAY38465, AAY38466, AAY38467, AAY38468,
DR AAY38469, AAY38470, AAY38471.
XX
XX New isolated human genes and the secreted polypeptides they encode
PT
XX Claim 1; Page 160; 227pp; English.
PS
XX This sequence represents a nucleic acid molecule which encodes a

CC secreted human protein. The gene number is given in the descriptor line.
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. AA06210) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 36 novel genes and their fragments (nucleic
CC acid sequences: AA06219-206263; amino acid sequences AA06219-206263)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 36
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AA06219 for described uses).

XX SQ Sequence 1194 BP; 358 A; 291 C; 249 G; 296 T; 0 other;

Query Match 1.4%; Score 29; DB 20; Length 1194;
Best Local Similarity 100.0%; Pred. No. 0.00046;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 CAGATATCATTTGATGGCTATGGCAA 81
DB 478 CAGATATCATTTGATGGCTATGGCAA 506

RESULT 14

AA022672
ID AAD22672 standard; DNA; 38 BP.

AC AAD22672;

DT 26-FEB-2002 (first entry)

XX Human ceramidase cDNA cloning reverse PCR primer #2.

XX Human; ceramidase; nephrotropic; antipsoriatic; antitense therapy;
KW gene therapy; proliferative disorder; cancer; cardiovascular disease;
KW inflammation; neurodegenerative disorder; cytostatic; immunosuppressive;
KW neurotropic; signal transduction; breast cancer; autoimmune disorder;
KW Alzheimer's disease; growth deficiency; lesion; lupus nephritis;
KW glomerular disease; PCR primer; ss.

XX Homo sapiens.

OS WO200155410-A2.

XX 02-AUG-2001.

XX 29-JAN-2001; 2001WO-US02866.

XX 28-JAN-2000; 2000US-178975P.

XX (MUSC-) MUSC FOUND RES DEV.

XX Hannun YA, El Bawab S;

DR WPI; 2002-025687/03.

XX Human mitochondrial ceramidase protein and gene, modulation of which is
PT useful for preventing and treating proliferative disorders e.g. cancer,
PT cardiovascular disease, inflammation and neurodegenerative disorders -
XX Example: Page 82; 113pp; English.

CC The present sequence is a PCR primer used for cloning human mitochondrial
CC ceramidase DNA. Ceramidase protein and gene are useful for treating a
CC disease or disorder involving cell over proliferation or sphingolipid
CC signal transduction especially breast cancer, cardiovascular disorder or
CC inflammation. Ceramidase protein and gene are useful for treating
CC disorders involving deficient cell proliferation or growth e.g.
CC neurodegenerative disorders (Alzheimer's disease), growth deficiencies
CC and lesions. Ceramidase protein is also useful for diagnosis of
CC hyperproliferative diseases. Ceramidase gene can be used as an immunogen

CC to generate antibodies which are useful for diagnosis, prevention and
CC treatment of hyperproliferative diseases and for detecting ceramidase
CC gene product in a biological sample. The hyperproliferative disorders
CC include cancers and autoimmune disorders such as lupus nephritis,
CC glomerular disease. Ceramidase gene is also useful in antitense therapy
CC and gene therapy.

XX SQ Sequence 38 BP; 11 A; 10 C; 10 G; 7 T; 0 other;

Query Match 1.2%; Score 25; DB 24; Length 38;
Best Local Similarity 100.0%; Pred. No. 0.05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1682 TGCAGCCAGCAAAACCTGATACAG 1706
DB 14 TGCAGCCAGCAAAACCTGATACAG 38

AA097636
ID AAA97636 standard; DNA; 24 BP.

AC AAA97636;

DT 02-FEB-2001 (first entry)

XX Mouse neutral/alkaline ceramidase PCR primer. SEQ ID NO:17.

XX Neutral/alkaline ceramidase; mouse; murine; recombinant production;
KW cellular ceramidase content control; antibody; membrane synthesis;
KW lipid engineering; ceramidase metabolism; drug development; PCR primer; ss.

OS Mus sp.

PN WO200058448-A1.

XX 05-OCT-2000.

XX 24-MAR-2000; 2000WO-JP01802.

XX 26-MAR-1999; 99JP-0084743.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Ito M;

DR WPI; 2000-619079/59.

XX Mammalian neutral/alkaline ceramidase applicable as lipid engineering
PT reagent for studying structure and functions of ceramidase as well as
PT developing drugs for diseases associated with ceramidase metabolism -
XX Example 7; Page 70; 76pp; Japanese.

XX The invention relates to a mouse neutral/alkaline ceramidase and to
CC nucleic acids encoding it. The invention also relates to expression
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
CC ceramidase, the recombinant production of the ceramidase, an antibody
CC against the ceramidase, and a method for controlling the ceramidase
CC content in cells and/or tissues by introducing the gene or its antisense
CC nucleic acid into the cells and/or tissues. The ceramidase can be used
CC as a lipid engineering reagent for studying structure and functions of
CC ceramidase as well as developing drugs for diseases associated with
CC ceramidase metabolism. The present sequence represents a PCR primer
CC used to amplify nucleic acids encoding mouse neutral/alkaline
CC ceramidase.

SO Sequence 24 BP; 8 A; 4 C; 6 G; 6 T; 0 other;

Query Match 1.2%; Score 24; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 885 GTTTGAGACACACATTATAGG 908
 Db 1 GTTTGAGACACACATTATAGG 24

RESULT 16
 ID AAA97637 standard; DNA: 24 BP.

AAA97637;

02-FEB-2001 (first entry)

Mouse neutral/alkaline ceramidase PCR primer, SEQ ID NO:18.

Neutral/alkaline ceramidase; mouse; murine; recombinant production;
 cellular ceramide content control; antibody; membrane synthesis;
 lipid engineering; ceramide metabolism; drug development; PCR primer; ss.

Mus sp.

WO200058448-A1.

05-OCT-2000.

24-MAR-2000; 2000MO-JP01802.

26-MAR-1999; 99JP-0084743.

(TAKI) TAKARA SHUZO CO LTD.

Ito M;

WPI; 2000-619079/59.

Mammalian neutral/alkaline ceramidase applicable as lipid engineering
 reagent for studying structure and functions of ceramide as well as
 developing drugs for diseases associated with ceramide metabolism.

Example 7; Page 70; 76pp; Japanese.

The invention relates to a mouse neutral/alkaline ceramidase and to
 nucleic acids encoding it. The invention also relates to expression
 vectors and host cells comprising DNA encoding mouse neutral/alkaline
 ceramidase, the recombinant production of the ceramidase, an antibody
 against the ceramidase, and a method for controlling the ceramide
 content in cells and/or tissues by introducing the gene or its antisense
 nucleic acid into the cells and/or tissues. The ceramidase can be used
 as a lipid engineering reagent for studying structure and functions of
 ceramide as well as developing drugs for diseases associated with
 ceramide metabolism. The present sequence represents a PCR primer
 used to amplify nucleic acids encoding mouse neutral/alkaline
 ceramidase.

Sequence 24 BP; 8 A; 7 C; 4 G; 5 T; 0 other;

Query Match 1.28; Score 24; DB 21; Length 24;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1089 TGATGAGTTGGGGCTCAATAT 1112
 Db 24 TGATGAGTTGGGGCTCAATAT 1

RESULT 17

ID AAV41320 standard; DNA: 4020 BP.

AAV41320;

09-NOV-1998 (first entry)

DE Human angiotensin converting enzyme gene.
 XX Angiotensin converting enzyme; ACE; hypertension; exercise; human;
 KM genetic marker; ss.

OS Homo sapiens.

Key Location/Qualifiers
 CDS 23..3943
 FT /*tag= a

FT sig-peptide 23..109

FT mat-peptide 110..3940

FT /*tag= b

FT /*tag= c

MO9831835-A1.

23-JUL-1998.

22-DEC-1997; 97MO-US22974.

27-MAY-1997; 97US-0048309.

16-JAN-1997; 97US-0035382.

(UYMA-) UNIV MARYLAND BALTIMORE.

(UYPI-) UNIV PITTSBURGH.

Ferrell RE, Hagberg JM;

WPI; 1998-414128/35.

P-P-SDB; AAM68155.

Analysis of genetic markers to identify subjects who will benefit

from exercise - also assessing risk of cardiovascular disease from

angiotensin-converting enzyme genotype

Disclosure; Page 35-41; 61pp; English.

This is the human angiotensin converting enzyme (ACE) gene. The

gene is polymorphic with 2 common alleles (I and D), resulting in

3 genotypes, II, ID and DD. It is an object of the invention to

identify individuals possessing a certain genotype and associated

with a certain phenotype. A claimed method comprises

improving by altering behavior. A claimed method comprises

identifying individuals having a certain phenotype, determining

the presence or absence of genetic markers associated with the

phenotype, and instituting a lifestyle change to exploit or

counteract the phenotype expressed by the gene marker. If the

phenotype is hypertension, the gene marker is at least one

insertion (I) ACE allele and exercise training is instituted to

decrease systolic and diastolic blood pressure. The gene marker

can be identified by PCR amplification (see AAV41321-22) of the

appropriate gene region. The general method can be used to

identify subjects who will benefit most from physical exercise

and also to identify those who are likely to be successful in

Sequence 4020 BP; 860 A; 1265 C; 1169 G; 726 T; 0 other;

Query Match 1.18; Score 22; DB 19; Length 4020;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 922 CAGAGGCCAAGAGCTGTATG 943
 Db 317 CAGAGGCCAAGAGCTGTATG 338

RESULT 18

ID AAA38330 standard; DNA: 4020 BP.

AAA38330;

AA38330;

XX 21-AUG-2000 (first entry)
 DT Human angiotensin-converting enzyme (ACE) coding region.
 XX
 DE Angiotensin-converting enzyme gene; ACE; coding region; polymorphism;
 KW polymorphic marker; cardiovascular disease; myocardial infarction;
 KM unstable angina; hypertension; atherosclerosis; stroke; prognosis;
 XX drug screening; treatment outcome; human; ds.
 XX Homo sapiens.
 OS
 XX WO200022166-A2.
 XX
 XX 20-APR-2000.
 XX
 XX 13-OCT-1999; 99WO-IB01678.
 XX
 XX 14-OCT-1998; 98US-0104286.
 PR 14-OCT-1998; 98US-0104302.
 XX
 XX (EURO-) EURONA MEDICAL AB.
 PA
 XX Norberg LT, Andersson MK, Lindstrom PHR, Jonsson L;
 PI
 XX WPI; 2000-318010/27.
 DR
 XX
 XX Assessing cardiovascular status in humans involves comparing test
 PT polymorphic pattern comprising polymorphic positions within genes
 PT encoding specific proteins, with reference polymorphic pattern
 XX
 XX Disclosure: Page 114-115; 126pp; English.
 XX
 XX The invention relates to a novel method of assessing the cardiovascular
 CC status in an individual and to newly identified polymorphisms in the
 CC genes encoding angiotensin-converting enzyme (ACE), angiotensin II
 CC receptor type 1 (AT1) and type 2 (AT2), angiotensinogen (AGT), renin,
 CC aldosterone synthase, endothelin receptor type A and beta-adrenergic
 CC receptors 1 and 2. The method comprises determining the sequence at one
 CC or more polymorphic positions within these genes, and comparing the
 CC pattern of polymorphisms from the individual with a reference polymorphic
 CC pattern obtained from a population of individuals exhibiting a
 CC predetermined cardiovascular disease status. The polymorphic markers are
 CC useful for determining the predisposition of an individual to
 CC cardiovascular disorders such as myocardial infarction, unstable angina,
 CC hypertension, atherosclerosis and stroke. They are also useful for
 CC predicting the likely cardiovascular status of a patient given a
 CC treatment regimen comprising administration of cardiovascular drugs
 CC (e.g., ACE inhibitors, beta-adrenergic receptor antagonists (beta-
 CC blockers) or calcium channel blockers). One or more polymorphic markers
 CC provides a basis for predicting the outcome of a treatment regimen.
 CC Fragments of the genes comprising a polymorphic site may be used as
 CC primers and probes for detecting genetic polymorphisms or in molecular
 CC library arrays for high throughput screening. The genes, and the proteins
 CC they encode are useful in the screening of potential cardiovascular
 CC drugs. Determination of an individual's polymorphic pattern reduces or
 CC eliminates trial and error in selecting a treatment for a particular
 CC individual cardiovascular patient. It also provides the ability to
 CC eliminate patients from clinical trials who are predicted to be
 CC non-responsive, or at a risk for an adverse response, to a particular
 CC treatment regimen. Adverse results in an early trial can be evaluated to
 CC identify polymorphic patterns so that the adverse results can be
 CC correlated with a sub-population of the test population, permitting
 CC exclusion of such sub-populations from the treatment group. Beneficial
 CC drugs can be approved for use in the appropriate population, thereby
 CC decreasing the number of patients required for a clinical trial, which in
 CC turn decreases the duration and cost of such trials. Sequences A38328 and
 CC A38330 represent, respectively, intron 16 and the coding region of
 CC the human ACE gene (GenBank X62855, J04144). The polymorphic sites
 CC identified are 375A/C, 582C/T, 731A/G, 1060G/A, 1215C/T, 2193G/A,
 CC 2228A/G, 2741G/T, 3132C/T, 3387T/C, 3503G/C, 3906G/A, and a deletion of
 CC nucleotides 1451-1783 in intron 16.
 CC
 XX

SQ Sequence 4022 BP; 857 A; 1261 C; 1174 G; 728 T; 0 other;
 Query Match 1.1%; Score 22; DB 21; Length 4020;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 922 CAGAGGCCAAGAGCTGTATG 943
 Db 317 CAGAGGCCAAGAGCTGTATG 338
 RESULT 19
 AAK9395
 ID AAK9395 standard; DNA: 4022 BP.
 XX
 AC AAK9395;
 XX
 DT 27-JUN-2002 (first entry)
 XX
 XX DNA of APP related human homologue hCP51674.
 DE
 XX Neuroprotective; nootropic; transgenic fly; Alzheimer's disease; Abeta;
 KW amyloid precursor protein; tissue-specific expression control; human APP;
 KW APP pathway modulator; gene therapy; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 148..3945
 FT /*tag= a
 FT /product= "Protein of human homologue hCP51674"
 FT /note= "No start codon"
 XX
 XX WO200226820-A2.
 XX
 XX 04-APR-2002.
 XX
 XX 01-OCT-2001; 2001WO-EP11345.
 XX
 XX 29-SEP-2000; 2000US-236893P.
 PR 14-JUN-2001; 2001US-298309P.
 XX
 XX (NOVS) NOVARTIS AG.
 PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 XX
 XX Cohen D, Dengler UJ, Finelli AL, Freuler F, Konsolaki M;
 PI Reinhardt MMH, Zisman S;
 XX
 XX WPI; 2002-315796/35.
 DR P-PSDB; AAO20501.
 XX
 XX New transgenic fly, containing DNA encoding an Abeta portion of human
 PT APP, useful for identifying agents which modulate the APP pathway and
 PT which can be used to treat Alzheimer's disease
 XX
 PS Example 4; Page 93-94; 129pp; English.
 XX
 XX The invention relates to a transgenic fly whose genome comprises DNA
 CC encoding a polypeptide having the Abeta portion of human amyloid
 CC precursor protein (APP), fused to a signal sequence. The DNA sequence
 CC encodes a 123 (Abeta40) or 129 (Abeta42) amino acid sequence, given in
 CC the specification. The DNA sequence is operably linked to a tissue-
 CC specific expression control sequence. Expression of the sequence gives
 CC the fly an altered phenotype. The purpose of the invention is for
 CC identifying agents that inhibit or promote the expression and/or function
 CC of genes or encoded polypeptides which modify the APP pathway. The agent
 CC is a compound, triple helix DNA, antisense oligonucleotide, double
 CC stranded RNA molecule, ribozyme, or particularly an antibody. It is used
 CC to treat conditions such as Alzheimer's disease. The agent can be used as
 CC an APP pathway modulator or in gene therapy. This polynucleotide sequence
 CC represents the DNA of the APP related human homologue hCP51674.
 CC
 XX Sequence 4022 BP; 856 A; 1262 C; 1174 G; 728 T; 2 other;
 XX

Query Match 1.1%; Score 22; DB 24; Length 4022;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 922 CAGAGGCCAAGAGCTGTATG 943
DB 319 CAGAGGCCAAGAGCTGTATG 340

RESULT 20
AA004027 ID AA004027 standard; CDNA; 4024 BP.

XX AC AA004027;
XX
XX 07-SEP-1990 (first entry)
XX
XX Human angiotensin converting enzyme (ACE) gene.
XX
XX human angiotensin converting enzyme; hypertension;
XX bradykinin;
XX synthetic.
XX
XX Key Location/Qualifiers
XX FT 23..51
XX FT sig_peptide /tag= a
XX FT CDS 23..3944
XX FT /*tag= b

XX PN WO9003435-A.
XX PD 05-APR-1990.
XX
XX 29-SEP-1989; 89WO-FR00469.
XX
XX 27-SEP-1988; 88FR-0012620.
XX
XX (INRM) INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE.
XX
XX Soudrier F, Alhenc-Gelas F, Hubert C, Corvol P;
XX
XX WPI: 1990-058128/08.
XX DR P-PSDB; AAR04111.
XX
XX New DNA encoding human angiotensin converting enzyme used eg in
XX diagnosis of hypertension, evaluation of enzyme inhibitors
XX
XX
XX Disclosure; ; P; French.
XX
XX This sequence encodes pre-ACE. In the mature protein the 29 amino acid
XX signal peptide is absent. ACE hydrolyses angiotensin I and/or kinins,
XX most notably bradykinin. Vectors contg. the given sequence are used to
XX produce ACE for therapeutic use eg to hydrolyse kinins implicated in
XX inflammation.
XX
XX Sequence 4024 BP; 857 A; 1263 C; 1175 G; 729 T; 0 other;

Query Match 1.1%; Score 22; DB 11; Length 4024;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 922 CAGAGGCCAAGAGCTGTATG 943
DB 317 CAGAGGCCAAGAGCTGTATG 338

RESULT 21
AA004027 ID AA004027 standard; CDNA; 4024 BP.
XX
XX AAX35850
XX
XX AAX35850;

XX DT 14-JUL-1999 (first entry)
XX
XX Human angiotensin converting enzyme DNA.
XX
XX Antisense oligonucleotide; mammalian; angiotensin converting enzyme;
XX ACE; hypertension; ss.
XX
XX Homo sapiens.
XX
XX PN WO9915643-A2.
XX PD 01-APR-1999.
XX
XX 25-SEP-1998; 98WO-US20121.
XX PF 25-SEP-1997; 97US-0059661.
XX PR 25-SEP-1997; 97US-0059661.
XX
XX (UWFL) UNIV FLORIDA.
XX
XX PI Mohuczy D, Phillips MI;
XX
XX DR WPI: 1999-276982/23.
XX
XX New antisense oligonucleotide compositions
XX
XX Disclosure; Page 142-143; 145pp; English.
XX
XX PS The present sequence represents human angiotensin converting enzyme
XX CC (ACE) DNA. The specification describes antisense oligonucleotides
XX CC (AAX35833-49) for mammalian angiotensin converting enzyme (ACE) mRNA.
XX CC The antisense oligonucleotides can be used for treating or preventing
XX CC hypertension. The antisense oligonucleotides can also be
XX CC labeled and used for visualising ACE mRNA in a cell, and for
XX CC producing transgenic animals.
XX
XX
XX Sequence 4024 BP; 857 A; 1263 C; 1175 G; 729 T; 0 other;

Query Match 1.1%; Score 22; DB 20; Length 4024;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 922 CAGAGGCCAAGAGCTGTATG 943
DB 317 CAGAGGCCAAGAGCTGTATG 338

RESULT 22
AAH57430 ID AAH57430 standard; CDNA; 5005 BP.

XX AC AAH57430;
XX
XX 10-SEP-2001 (first entry)
XX
XX Human intestine cell specific CDNA sequence SEQ ID NO: 270.
XX
XX Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
XX lung; liver; uterus; ovary; stomach; kidney; pancreas; ss;
XX metabolic disease; developmental disease; cytostatic; immunomodulatory;
XX neuroprotective; gene therapy; cancer; Immunopathology; neuropathology.
XX
XX Homo sapiens.
XX
XX PN WO200132927-A2.
XX PD 10-MAY-2001.
XX
XX 02-NOV-2000; 2000WO-US30396.
XX PF 04-NOV-1999; 99US-0163508.
XX PR (INCY-) INCYTE GENOMICS INC.
XX PA

XX Sornasse T, Sellhamer JJ, Watson GA;
XX WPI: 2001-291057/30.
XX
XX
XX New cell and tissue specific polynucleotides useful for diagnosis,
XX prognosis or monitoring of treatments for disorders where the gene is
XX associated with a cancer, immunopathology or neuropathology -
XX
XX Claim 1: Page 197-198; 327pp; English.
XX
XX AAH57161 to AAH5756 represent cell and tissue specific polynucleotide
XX sequences (I). (I) can have cytosstatic, immunomodulatory and
XX neuroprotective activities, and can be used in gene therapy. (I) and
XX proteins (II) encoded by then are used in high throughput screening
XX assays to select DNA molecules, RNA molecules, peptide nucleic acids,
XX mimetics, peptides, proteins, agonists, antagonists, antibodies or
XX their fragments, immunoglobulins, inhibitors, drug compounds and
XX pharmaceutical agents. Expression of (I) in a sample indicates the
XX differentiation of embryonic stem cells into a tissue selected from
XX brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
XX tissues. (I) and (II) are used to produce an expression profile that
XX defines a metabolic or developmental process, treatment, condition,
XX disease or disorder. The gene profile can be used for diagnosis,
XX prognosis or monitoring of treatments and for investigating a
XX predisposition to a disorder where the gene is associated with a
XX cancer, immunopathology or neuropathology.
XX
XX Sequence 5005 BP; 1063 A; 1584 C; 1420 G; 936 T; 2 other;
SQ
Query Match 1.1%; Score 22; DB 22; Length 5005;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 922 CAGAAAGCCAGAGAGCTGTATG 943
DB 316 CAGAAAGCCAGAGAGCTGTATG 337
RESULT 23
AAK69978
ID AAK69978 standard; DNA: 21 BP.
XX
XX AAK69978;
XX
XX 20-JUL-2001 (first entry)
XX
XX PCR primer used to isolate rat cDNA.
XX
XX Rat; acid amide bond hydrolysis; ceramide; cytosstatic; cancer;
XX PCR primer; ss.
XX
XX Rattus norvegicus.
XX
XX JP2001057890-A.
XX
XX 06-MAR-2001.
XX
XX 23-AUG-1999; 99JP-0235218.
XX
XX 23-AUG-1999; 99JP-0235218.
XX
XX 23-AUG-1999; 99JP-0235218.
XX
XX (MITU) MITSUBISHI CHEM CORP.
XX
XX WPI: 2001-304133/32.
XX
XX Novel protein, used to treat abnormal cell growth e.g. in cancers, is
XX capable of hydrolyzing the acid amide bond between sphingosine and the
XX fatty acid in ceramide -
XX
XX Example 3: Page 14; 14pp; Japanese.
XX
XX The present sequence is provided in a specification relating to a

CC protein having a 761 residue amino acid sequence or its variant
CC comprising a replacement, deletion, insertion, addition or reversion of
CC at least one amino acid, but retaining the activity of hydrolyzing the
CC acid amide bond between sphingosine and the fatty acid in a ceramide.
CC The protein can be used to treat diseases caused by abnormal cell
CC growth such as cancers.
XX
XX Sequence 21 BP; 6 A; 1 C; 8 G; 6 T; 0 other;
SQ
Query Match 1.0%; Score 21; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 562 CTGCTGATTTGATGAGAA 582
DB 1 CTGCTGATTTGATGAGAA 21
RESULT 24
AAK35851
ID AAK35851 standard; DNA: 3942 BP.
XX
XX AAK35851;
XX
XX 14-JUL-1999 (first entry)
XX
XX Rat angiotensin converting enzyme DNA.
XX
XX Antisense oligonucleotide; mammalian; angiotensin converting enzyme;
XX ACE; hypertension; ss.
XX
XX Rattus norvegicus.
XX
XX WO9915643-A2.
XX
XX 01-APR-1999.
XX
XX 25-SEP-1998; 98WO-US20121.
XX
XX 25-SEP-1997; 97US-0059661.
XX
XX (UYFL) UNITV FLORIDA.
XX
XX Mohuczy D, Phillips MI;
XX
XX WPI: 1999-276982/23.
XX
XX New antisense oligonucleotide compositions
XX
XX Disclosure: Page 143-144; 145pp; English.
XX
XX The present sequence represents rat angiotensin converting enzyme
XX (ACE) DNA. The specification describes antisense oligonucleotides
XX (AAK35833-49) for mammalian angiotensin converting enzyme (ACE) mRNA.
XX The antisense oligonucleotides can be used for treating or preventing
XX hypertension. The antisense oligonucleotides can also be
XX labeled and used for visualizing ACE mRNA in a cell, and for
XX producing transgenic animals.
XX
XX Sequence 3942 BP; 939 A; 1124 C; 1061 G; 818 T; 0 other;
SQ
Query Match 1.0%; Score 21; DB 20; Length 3942;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 923 AGAAGGCCAGAGAGCTGTATG 943
DB 314 AGAAGGCCAGAGAGCTGTATG 334
RESULT 25
AAK68974/C
ID AAK68974 standard; DNA: 5681 BP.

XX AAK68974;
AC 06-NOV-2001 (first entry)
XX
DT
XX
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:23786.
XX
KW Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX WO200157182-A2.
PN
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0215135.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226886.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228624.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234297.
PR 25-SEP-2000; 2000US-0234498.
PR 26-SEP-2000; 2000US-0234584.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236602.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.


```

PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 23786; 3071bp + Sequence Listing; English.
XX
CC AA54951 to AA64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AA62170 to AA61921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AA64703
CC to AA67694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AA54942 to AA54950 and AA62169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 5681 BP; 1206 A; 1760 C; 1554 G; 1161 T; 0 other;

```

```

Query Match 1.0%; Score 20; DB 22; Length 5681;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1575 CAACATGAGCAGTGTCCCG 1594
   |||||||||||||||||||||
Db 3580 CAACATGAGCAGTGTCCCG 3561

```

Search completed: July 5, 2003, 13:28:47
Job time : 468 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

SUMMARIES

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 11:30:39 ; Search time 539f. Seconds

(without alignments)
11061.334 Million cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttccagtgcctactactggtg.....ttgaagtgtcactactg 2049

Scoring table: OLIGO_NUC

Searched: 2054640 seqs, 14551402878 residues

Word size : 20

Total number of hits satisfying chosen parameters: 119

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

Genembl:*

1: gb_ba:*

2: gb_hgt:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_pro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hgt_hum:*

31: em_hgt_inv:*

32: em_hgt_other:*

33: em_hgt_mus:*

34: em_hgt_pln:*

35: em_hgt_rod:*

36: em_hgt_mam:*

37: em_hgt_vrt:*

38: em_sy:*

39: em_hgt_hum:*

40: em_hgt_mus:*

41: em_hgt_other:*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	2049	100.0	3108	10	AB037111	AB037111 Mus muscu
2	2016	98.4	4835	10	AB037811	AB037811 Mus muscu
3	1493	72.9	1571	10	BC022604	BC022604 Mus muscu
4	192	9.4	165407	2	AC117816	AC117816 Mus muscu
5	179	8.7	165407	2	AC117816	AC117816 Mus muscu
6	179	8.7	184101	2	AC113485	AC113485 Mus muscu
7	72	3.5	184101	2	AC113485	AC113485 Mus muscu
8	68	3.3	2283	6	E50433	E50433 Novel prote
9	68	3.3	2456	10	AB057433	AB057433 Rattus no
10	53	2.6	539	6	E50435	E50435 Novel prote
11	47	2.3	235	9	HUMD4C12M5	AF250847 Homo sapi
12	47	2.3	2289	9	AF250847	AF250847 Homo sapi
13	47	2.3	2507	6	AX207121	AX207121 Sequence
14	47	2.3	162228	2	AL450382	AL450382 Human DNA
15	47	2.3	176865	2	AL513306	AL513306 Homo sapi
16	47	2.3	178899	2	AL589794	AL589794 Human DNA
17	39	1.9	110000	2	AC055726_2	Continuation (3 of
18	36	1.8	105838	2	AC105714	AC105714 Rattus no
19	32	1.6	105838	2	AC105714	AC105714 Rattus no
20	30	1.5	187130	2	AL583858	AL583858 Homo sapi
21	25	1.2	38	6	AX207126	AX207126 Sequence
22	23	1.1	164805	8	OSJN00174	AF662970 Oryza sat
23	23	1.1	175330	2	AC012131	AC012131 Homo sapi
24	22	1.1	168	9	FI93462802	FI93463 Pan trogl
25	22	1.1	4020	6	ARI37383	ARI37383 Sequence
26	22	1.1	4020	6	HUMAIICEB	JO4144 Human angio
27	22	1.1	4022	6	AX429555	AX429555 Sequence
28	22	1.1	4024	6	A00914	A00914 H. sapiens g
29	22	1.1	24070	9	AF118569	AF118569 Homo sapi
30	22	1.1	80858	9	AC007360	AC007360 Homo sapi
31	22	1.1	201500	2	AC103946	AC103946 Mus muscu
32	21	1.0	21	6	E50434	E50434 Novel prote
33	21	1.0	4014	10	AF201331	AF201331 Rattus no
34	21	1.0	4014	10	AF201332	AF201332 Rattus no
35	21	1.0	4014	10	RND03708	U03708 Rattus norv
36	21	1.0	4142	10	RND03734	U03734 Rattus norv
37	21	1.0	53423	9	AL591419	AL591419 Human DNA
38	21	1.0	76700	2	AC131469	AC131469 Rattus no
39	21	1.0	100858	2	AC096435	AC096435 Rattus no
40	21	1.0	179743	2	AC098290	AC098290 Rattus no
41	21	1.0	181663	2	AC068982	AC068982 Homo sapi
42	21	1.0	186651	2	AC120772	AC120772 Rattus no
43	21	1.0	191192	2	AC106294	AC106294 Rattus no
44	21	1.0	197174	2	AC094442	AC094442 Rattus no
45	21	1.0	212255	2	AC096705	AC096705 Rattus no
46	20	1.0	750	11	AL592564	AL592564 Meleagris
47	20	1.0	1065	5	HSM800587	AL080086 Homo sapi
48	20	1.0	1680	5	AF231711	AF231711 Xenopus l
49	20	1.0	17411	1	MYCRN1	L35043 Mycoplasma
50	20	1.0	23461	2	AC116980	AC116980 Dictyoste
51	20	1.0	43928	2	AC090263	AC090263 Homo sapi
52	20	1.0	53772	2	AC097284	AC097284 Rattus no
53	20	1.0	60657	2	AC101392	AC101392 Mus muscu
54	20	1.0	71669	2	AC102279	AC102279 Mus muscu
55	20	1.0	90820	2	AC111155	AC111155 Homo sapi
56	20	1.0	97143	9	AC025756	AC025756 Homo sapi
57	20	1.0	97488	8	ATT29H11	AL049659 Arabidops
58	20	1.0	110000	10	AE014173_1	Continuation (2 of
59	20	1.0	118873	9	HS1184F4	AL034450 Human DNA
60	20	1.0	141418	9	AC011741	AC011741 Homo sapi
61	20	1.0	142500	2	AC117145	AC117145 Rattus no
62	20	1.0	144091	8	AP003022	AP003022 Oryza sat
63	20	1.0	144269	2	AC098342	AC098342 Rattus no
64	20	1.0	144734	2	AC123948	AC123948 Mus muscu
65	20	1.0	145945	2	AC119369	AC119369 Rattus no

C	66	20	1.0	148526	10	AL732503	AL732503	Mouse	DNA
C	67	20	1.0	149768	2	AC016239	AC016239	Homo	sapi
C	68	20	1.0	150683	2	AC120517	AC120517	pan	trog
C	69	20	1.0	151371	2	AC091889	AC091889	Homo	sapi
C	70	20	1.0	154373	10	AC121988	AC121988	Mus	muscu
C	71	20	1.0	157821	2	AC120725	AC120725	Rattus	no
C	72	20	1.0	159283	2	AC107705	AC107705	Mus	muscu
C	73	20	1.0	161739	9	AC105277	AC105277	Homo	sapi
C	74	20	1.0	163040	2	AC112470	AC112470	Rattus	no
C	75	20	1.0	163312	9	AC007955	AC007955	Homo	sapi
C	76	20	1.0	165614	2	AC125707	AC125707	Rattus	no
C	77	20	1.0	167340	2	AC111314	AC111314	Rattus	no
C	78	20	1.0	167658	2	AC123942	AC123942	Mus	muscu
C	79	20	1.0	168440	2	AC109112	AC109112	Rattus	no
C	80	20	1.0	170325	2	AC109013	AC109013	Rattus	no
C	81	20	1.0	173840	9	AC092445	AC092445	Homo	sapi
C	82	20	1.0	174711	2	AC025515	AC025515	Homo	sapi
C	83	20	1.0	175080	2	AC113545	AC113545	Rattus	no
C	84	20	1.0	176951	2	AL845340	AL845340	Mus	muscu
C	85	20	1.0	177172	9	CNS00006	AL049832	Human	chr
C	86	20	1.0	178298	2	AC094759	AC094759	Rattus	no
C	87	20	1.0	178652	10	AL713853	AL713853	Mouse	DNA
C	88	20	1.0	180682	2	AC109776	AC109776	Rattus	no
C	89	20	1.0	182199	2	AC095097	AC095097	Rattus	no
C	90	20	1.0	185306	2	AC013570	AC013570	Homo	sapi
C	91	20	1.0	185761	2	AC108814	AC108814	Mus	muscu
C	92	20	1.0	186314	10	AC003694	AC003694	Mus	muscu
C	93	20	1.0	187507	2	AC124724	AC124724	Mus	sapi
C	94	20	1.0	188061	2	AC007373	AC007373	Homo	sapi
C	95	20	1.0	188851	10	AL591905	AL591905	Mouse	DNA
C	96	20	1.0	188921	9	AC096730	AC096730	Homo	sapi
C	97	20	1.0	189891	2	AC119058	AC119058	Papio	cyn
C	98	20	1.0	192802	2	AC103600	AC103600	Mus	muscu
C	99	20	1.0	193865	2	AC122270	AC122270	Mus	muscu
C	100	20	1.0	196664	2	AC131185	AC131185	Mus	muscu
C	101	20	1.0	198954	2	AC099552	AC099552	Homo	sapi
C	102	20	1.0	200065	10	AC080021	AC080021	Genomic	s
C	103	20	1.0	201839	2	AC098333	AC098333	Rattus	no
C	104	20	1.0	202066	2	AC110633	AC110633	Rattus	no
C	105	20	1.0	202249	2	AC095427	AC095427	Rattus	no
C	106	20	1.0	206819	2	AC016029	AC016029	Homo	sapi
C	107	20	1.0	207538	2	AC106886	AC106886	Homo	sapi
C	108	20	1.0	209171	2	AL845279	AL845279	Mus	muscu
C	109	20	1.0	210321	2	AC114001	AC114001	Mus	muscu
C	110	20	1.0	213473	2	AC094371	AC094371	Rattus	no
C	111	20	1.0	213901	2	AC093472	AC093472	Mus	muscu
C	112	20	1.0	216238	10	AL513356	AL513356	Mouse	DNA
C	113	20	1.0	218553	2	AC079940	AC079940	Mus	muscu
C	114	20	1.0	221712	2	AC126447	AC126447	Mus	muscu
C	115	20	1.0	223368	2	AC103652	AC103652	Mus	muscu
C	116	20	1.0	223426	2	AC121074	AC121074	Canis	fam
C	117	20	1.0	240233	2	AC127342	AC127342	Mus	muscu
C	118	20	1.0	258661	2	AC129078	AC129078	Mus	muscu
C	119	20	1.0	269423	2	AC129333	AC129333	Mus	muscu

ALIGNMENTS

RESULT	1
AB037111	
LOCUS	3108 bp mRNA linear ROD 08-JUN-2001
DEFINITION	Mus musculus LCDase mRNA for neutral ceramidase, complete cds.
ACCESSION	AB037111
VERSION	AB037111.1 GI:7576308
KEYWORDS	neutral ceramidase.
SOURCE	Mus musculus liver cDNA to mRNA.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (sites)
AUTHORS	Tani,M., Okino,N., Mori,K., Tanigawa,T., Ito,H. and Ito,M.
TITLE	Molecular cloning of the full-length cDNA encoding mouse neutral

neramidase. A novel but highly conserved gene family of neutral/alkaline ceramidases

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 3108)
Okino, N. and Ito, M.
Direct Submission
Submitted (14-JAN-2000) Makoto Ito, Kyushu University, Graduate School of Bioresource and Bioenvironmental Sciences, Department of Bioscience and Biotechnology; 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan (E-mail:makoto@agr.kyushu-u.ac.jp,
Tel.:81-92-642-2900, Fax:81-92-642-2900)

FEATURES
Source

1..3108
Location/Qualifiers

/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="Liver"
1..3108
/gene="Lcdase"
726..2996
/gene="Lcdase"
/codon_start=1
/product="neutral ceramidase"
/protein_id="BAA94545.1"
/db_xref="GI:7576309"
/translation="MAKRTSTLEAFLLVMTVTVALLTLLFTYSGTIEHKHDS
GNHFSTTGSTTTQPTPTQTTPNPFSPSRNSGYIGRADGCGQSDILMKGKK
GONRGLITRLFSRAFLIADGGNRNAFYVELCMIQSRLRKLESKYSLYFR
RDNYLSLIHTHSAGPFGYYTLILAEGSNRFQYIVGIMKSIDIAHTNLPGGKSLYF
IFIKGVANVOINRSPSSYLNPDSSEARASNTMDKMLVLVDLNGEDLGISW
AIIHGNNMSNHFPVSNMGYAALFDPEKKKGYLPQGPPVAGPASNDGVSPIILL
GPCVNTGESCDNDKSTCPNGSPSCMASGCGOMPESTHIIIGHIITYOKAKELVASAS
QEYGVPLVAHQVMNTDVSVOLNATHIVTKCPALGSFAAGTIDVGSLINTGTT
EGDFMDLTRDLQWKPSSEIVECOKRPILHSGELTIIPHPMPDIYDVQIVYGSI
AIIAIPELITMSGRRFEALIKKEFALGMDMNTVINLYTHYTTEEOAQOAG
YEASTISPHPTLSAYTOLFEDLKALATOVAMMSGSEPPEFNKLIASLIPNIAD
APIKHGHCVDPAPKPREYGEVEVETVGCANPKNSAENQHOFLEYEKEDSVAPR
OIMENDSWETRMYHKGILGELSNAITWIHLPTAIRPFIIRIFGRNKDELLPAPV
ILAEGLISSPEEVTT"

BASE COUNT 888 a 760 c 714 g 746 t

ORIGIN

Query Match 100.0%; Score 2049; DB 10; Length 3108;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TTCACTGGCTCTCATATTGGCTTGGAGACGATTCACAGACAAGTGTCAATATC 60
|||||
Db 948 TTCAGTGGCTACTACATATGGCTTGGAGACGATTCACAGACAAGTGTCAATATC 1007
|||||
Oy 61 AATTGATGGCTATGGCAAAAATGGCCAGAATGCACGGGGTCTCTCCACCAGGCTGTC 120
|||||
Db 1008 AATTGATGGCTATGGCAAAAATGGCCAGAATGCACGGGGTCTCTCCACCAGGCTGTC 1067
|||||
Oy 121 AGCGGTGTTTAATCTTGGCGGATTCAGATGGGTCAAATGCAATGGCATTTGTGAAGCTG 180
|||||
Db 1068 AGCGGTGTTTAATCTTGGCGGATTCAGATGGGTCAAATGCAATGGCATTTGTGAAGCTG 1127
|||||
Oy 181 GAACATATGATGATTTCCCAACGACTGAGTGGAGGTCGGAAGAGACTAGAGATGAA 240
|||||
Db 1128 GAACATATGATGATTTCCCAACGACTGAGTGGAGGTCGGAAGAGACTAGAGATGAA 1187
|||||
Oy 241 TATGAGTCTCTGTATCGAAGACAAATTTATCTCTGATGATGCCATTCACACACACTTGGC 300
|||||
Db 1188 TATGAGTCTCTGTATCGAAGACAAATTTATCTCTGATGATGCCATTCACACACACTTGGC 1247
|||||
Oy 301 CCACGAGGGTTTTTCCAATATACACTCTATATCTCGCCACGAGGAGATTGACAAACCGG 360
|||||
Db 1248 CCACGAGGGTTTTTCCAATATACACTCTATATCTCGCCACGAGGAGATTGACAAACCGG 1307
|||||
Oy 361 ACCTTTCAATACATATGCTCTGGGATCATGAAAGACATGATGATATAGCTACACAAATCTT 420
|||||
Db 1308 ACCTTTCAATACATATGCTCTGGGATCATGAAAGACATGATGATATAGCTACACAAATCTT 1367

OY	421	AAACGAGCAAAATCTTTATCACAAAGAAATGTGCTATGTGACATCAACCGAAG	480
Db	1368	AAACGAGCAAAATCTTTATCACAAAGAAATGTGCTATGTGACATCAACCGAAG	1427
OY	481	CCCTCCTCTTACCTCTTGCAATCCACAGTCACAGAGAGCAAGTATCTTCAACACAGAC	540
Db	1428	CCCTCCTCTTACCTCTTGCAATCCACAGTCACAGAGAGCAAGTATCTTCAACACAGAC	1487
OY	541	AAGCAATGCTGTGCTTGAAACTGGTGAATTTGAATGAGAAAGACTTGGCTTATCAGC	600
Db	1488	AAGCAATGCTGTGCTTGAAACTGGTGAATTTGAATGAGAAAGACTTGGCTTATCAGC	1547
OY	601	TGTTTGGCATCAACCCCGTGAGCATGAACAAATGCAACCACTTTGCAATAGAGCAAT	660
Db	1548	TGTTTGGCATCAACCCCGTGAGCATGAACAAATGCAACCACTTTGCAATAGAGCAAT	1607
OY	661	ATGGGCTATGGGGCTTACCTTTTGAGCAAAAAAAGAAACAAGCTATCTGCTGGACAG	720
Db	1608	ATGGGCTATGGGGCTTACCTTTTGAGCAAAAAAAGAAACAAGCTATCTGCTGGACAG	1667
OY	721	GGACCGTTTGTAGCAGGCTTTCCTTATCAAAATCTGGAGACGTCACCCACATTTCTT	780
Db	1668	GGACCGTTTGTAGCAGGCTTTCCTTATCAAAATCTGGAGACGTCACCCACATTTCTT	1727
OY	781	GGCCCGCATTTGTGTCACACAGAGGGGAGCTTGTGACAAACGACAGAGCACTGTCCAC	840
Db	1728	GGCCCGCATTTGTGTCACACAGAGGGGAGCTTGTGACAAACGACAGCACTGTCCAC	1787
OY	841	GGTGGGCTTACATGTGCTGATGGCCAGCGCACTTGACAAAGCATGTTTGAAGACACAC	900
Db	1788	GGTGGGCTTACATGTGCTGATGGCCAGCGCACTTGACAAAGCATGTTTGAAGACACAC	1847
OY	901	ATTATAGAGCGGATCATCTATACAGAGGCCAAGAGCTGTATTCCTTGCCCTCCAGAG	960
Db	1848	ATTATAGAGCGGATCATCTATACAGAGGCCAAGAGCTGTATTCCTTGCCCTCCAGAG	1907
OY	961	GTGACCGGCGCAGAGCTTGCGACGTCACACAGGGGTGAACAATGACAGATGTGAGGCTCC	1020
Db	1908	GTGACCGGCGCAGAGCTTGCGACGTCACACAGGGGTGAACAATGACAGATGTGAGGCTCC	1967
OY	1021	CTCAATGCCACACACACAGTGAAGACGTGTAAACCTGGCCCTGAGGCTACAGTTTGC	1080
Db	1968	CTCAATGCCACACACACAGTGAAGACGTGTAAACCTGGCCCTGAGGCTACAGTTTGC	2027
OY	1081	GGCAATATGTGAGATTTCGGGCTCAATATTACACAGGGAATACAGGAAGGGAATCA	1140
Db	2028	GGCAATATGTGAGATTTCGGGCTCAATATTACACAGGGAATACAGGAAGGGAATCA	2087
OY	1141	TTTGGGAGCATCTTTCGGGACCACTCTTTGGGAAAACCATGCAAGAAATGTGAGAGT	1200
Db	2088	TTTGGGAGCATCTTTCGGGACCACTCTTTGGGAAAACCATGCAAGAAATGTGAGAGT	2147
OY	1201	CAGAAACCAACCAATCTGCTTACAGAGTGAGAGCTGACGATACCAATCTTGGCAA	1260
Db	2148	CAGAAACCAACCAATCTGCTTACAGAGTGAGAGCTGACGATACCAATCTTGGCAA	2207
OY	1261	CCAATATATGTGATGTTCAATGTTTACCGTGTGGTCTTGGCCATATGCTATCCCT	1320
Db	2208	CCAATATATGTGATGTTCAATGTTTACCGTGTGGTCTTGGCCATATGCTATCCCT	2267
OY	1321	GGGCAATTAACAAACCATGTGGGAGCAAGATTTTGAGGAAATTAATAAAAAAGATTGCA	1380
Db	2268	GGGCAATTAACAAACCATGTGGGAGCAAGATTTTGAGGAAATTAATAAAAAAGATTGCA	2327
OY	1381	CTTATATGGGATGAAGATATGACCGTGTGTTATGCGAGGTCTATAGCAATGTTTATACAT	1440
Db	2328	CTTATATGGGATGAAGATATGACCGTGTGTTATGCGAGGTCTATAGCAATGTTTATACAT	2387
OY	1441	TACATTCCCATATATGAAGAAATACAGGCTACAGGGGTATACAGGAGCATCTACAAATCTAT	1500
Db	2388	TACATTCCCATATATGAAGAAATACAGGCTACAGGGGTATACAGGAGCATCTACAAATCTAT	2447

QY	1501	GGACCAACACCCGTGTGCAATACATCCAACTCTTGAGAGACCTTGGCAAGGCAATGCT	1560
Db	2448	GGACCAACACCCGTGTGCAATACATCCAACTCTTGAGAGACCTTGGCAAGGCAATGCT	2507
QY	1561	ACGCAACAGTAGGCAACATGAGCAGAGTGGTCCGAGCCTCCATTCCTCAAAAACCTATA	1620
Db	2508	ACGCAACAGTAGGCAACATGAGCAGAGTGGTCCGAGCCTCCATTCCTCAAAAACCTATA	2567
QY	1621	GCTTCACTTATTTCTATATATTGGCGATATGAGCACCAATTTGGCAACATTTTGGGATGTC	1680
Db	2568	GCTTCACTTATTTCTATATATTGGCGATATGAGCACCAATTTGGCAACATTTTGGGATGTC	2627
QY	1681	TTGCAAGCAGAGAAAACCTGAAATACAGATGGGAGAAAGTGTTGAAGTATATTTGTAGGC	1740
Db	2628	TTGCAAGCAGAGAAAACCTGAAATACAGATGGGAGAAAGTGTTGAAGTATATTTGTAGGC	2687
QY	1741	GCTAAACCAAGAATTCAGCAGAGAACAGACCATCAACCTTCCTACATGTGGAGAA	1800
Db	2688	GCTAAACCAAGAATTCAGCAGAGAACAGACCATCAACCTTCCTACATGTGGAGAA	2747
QY	1801	TACGAGACTCTGATGCTGACTGGCAGATAATGTATAACGATGCCCTCTGGGAGACGAGG	1860
Db	2748	TACGAGACTCTGATGCTGACTGGCAGATAATGTATAACGATGCCCTCTGGGAGACGAGG	2807
QY	1861	TTTATATGTGGCAAAAGGAAATACTGGGTGTGAGCAATGCACAAATATATACTGGCAATATCCA	1920
Db	2808	TTTATATGTGGCAAAAGGAAATACTGGGTGTGAGCAATGCACAAATATATACTGGCAATATCCA	2867
QY	1921	GATACGTGCTACCCCTGGAAATCTACAGAAATAGATATTTTGGACACCAATCGGAGACAGAA	1980
Db	2868	GATACGTGCTACCCCTGGAAATCTACAGAAATAGATATTTTGGACACCAATCGGAGACAGAA	2927
QY	1981	CTTCTGAAACCCGCTGTCTATCTAGCATTTTGAAGAAATTTCTTCCTTTTGAAGTTGTC	2040
Db	2928	CTTCTGAAACCCGCTGTCTATCTAGCATTTTGAAGAAATTTCTTCCTTTTGAAGTTGTC	2987
QY	2041	ACTACTTAG 2049	
Db	2988	ACTACTTAG 2996	

```

AB037181 RESULT 2
LOCUS AB037181
DEFINITION AB037181 4835 bp mRNA linear ROD 08-JUN-2001
ACCESSION Mus musculus BChase mRNA for neutral ceramidase, complete cds.
VERSION AB037181.1 GI:7576310
KEYWORDS neutral ceramidase.
SOURCE Mus musculus tissue_library:brain cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murineae; Mus.
1 (sites)
Tani,M., Okino,N., Mori,K., Tanigawa,T., Izu,H. and Ito,M.
Molecular cloning of the full-length cDNA encoding mouse neutral
ceramidase. A novel but highly conserved gene family of
neutral/alkaline ceramidases
J. Biol. Chem. 275 (15), 11229-11234 (2000)
20219171
2 (bases 1 to 4835)
Okino,N. and Ito,M.
Direct Submission
Submitted (17-JAN-2000) Makoto Ito, Kyushu University, Graduate
School of Bioresource and Bioenvironmental Sciences, Department of
Bioscience and Biotechnology; 6-10-1 Hakozaki, Higashi-ku, Fukuoka
812-8581, Japan (E-mail:makotoi@agr.kyushu-u.ac.jp,
Tel:81-92-642-2900, Fax:81-92-642-2900)
location/Qualifiers
1. 4835
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_library="Brain"
1. 4835
gene

```

```

CDS
    /gene="BCDase"
    /gene="BCDase"
    /codon_start=1
    /product="neutral ceramidase"
    /protein_id="BA94546.1"
    /db_xref="GI:7576311"
    /translation="MAKRTSTLEAFLEFLVIMTVIVALLTLLEVTSGTLENKDS
GNHMFSTLGSSTTQPPITQPNFSPFNSGYIGVRACRQGVSDINLMGKRN
GONARGLTRLESRAPILADPDGNSRMAVSVELCMIISQRLREVLKRESHSGSLR
RDNVLSIAHTSGPAGFQYTLTYLILASGFSNRTFOYIVSGIMKSIDAHNLMKRG
IFINKGNVANOINRSPSSYLINPQSEARYSNIDKEMLVKLDNEDJHILSMF
AHPVSNNNSNHPVNSDNNGYAAIYFPEOKNKGILPQSGPEYAGRASSLSLGEVSRIL
GPHCVNIGESCDNDSTCPNGSPSCMASPGQDMESTHIIIGRIILYQAKELVASAS
QEVTPVLAHQVMNTDIVSYLNATHVTKCKPALGYSFAIGTIDVSGILITVQGT
BEDPMDLRLDLGKPSSEIYEKCPKPIILHSELTPHMOPIIDVQIYVQSL
AIAAIPEGLTMSGRREAIKKEFALYKMDMTYVIALSVYTHYITTYEYOQR
YEASTITGVPHLISAVIOLEFRLAKAIADTVANSSGEPEPKULIASLIPNADR
APIGHFDVLOPAKREYRGVEVVIIPGANPKNSAENQHOTPLTVEKIEDSVADM
QIMYNDASHERTFYHKHGLISNMTIIMHPDIAIPGIYRIRLRFHNRKQELLPAPV
ILAFGISSPFEVVT"
BASE COUNT      1460 a      996 c     1048 g     1331 t
ORIGIN
Query Match      98.4%: Score 2016; DB 10; Length 4835;
Best Local Similarity 100.0%: Pred. No. 0;
Matches 2016; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    34 GATTGCACAGCAGCAAGTGTACATATCAATTTGATGGCTATGCAAAATGGCCAGAT 93
    |||||
    696 GATTGCACAGCAGCAAGTGTACATATCAATTTGATGGCTATGCAAAATGGCCAGAT 755
    |||||
    94 GCACGGGGCTCTCCTCACACAGGCTGTTACGCCCTGTTTATCTTGCGGATCCAGATGG 153
    |||||
    756 GCACGGGGCTCTCCTCACACAGGCTGTTACGCCCTGTTTATCTTGCGGATCCAGATGG 815
    |||||
    154 TCAAAATCGAATGGCAATTTGTGAGCGTGGAACTATGATGATTTCCCAACGACTGAGGCTG 213
    |||||
    816 TCAAAATCGAATGGCAATTTGTGAGCGTGGAACTATGATGATTTCCCAACGACTGAGGCTG 875
    |||||
    214 GAGGTCTCTGAAGAGACTAGAGAGTAAATATGGCTCTCTGTATCGAAGAGACAATGTATC 273
    |||||
    876 GAGGTCTCTGAAGAGACTAGAGAGTAAATATGGCTCTCTGTATCGAAGAGACAATGTATC 935
    |||||
    274 CTGAGTGCCTTACACACACACTCTGCCCCAGAGAGGTTTTCATATACACTCTATATA 333
    |||||
    936 CTGAGTGCCTTACACACACACTCTGCCCCAGAGAGGTTTTCATATACACTCTATATA 995
    |||||
    334 CTGGCAGAGGAGGATTCAGCAACCGGACCTTTCAGTACATAGCTCTGCGGATCANTGAG 393
    |||||
    996 CTGGCAGAGGAGGATTCAGCAACCGGACCTTTCAGTACATAGCTCTGCGGATCANTGAG 1055
    |||||
    394 AGCATTTGATATAGCTCACACAATCTTAAACAGGCAAAATCTTTATCAACAAGAAAT 453
    |||||
    1056 AGCATTTGATATAGCTCACACAATCTTAAACAGGCAAAATCTTTATCAACAAGAAAT 1115
    |||||
    454 GTTGCTAATGTCAAGTCAACCGAAAGCCCTCTCTTACCTTCTGTAATCCACAGTCAAG 513
    |||||
    1116 GTTGCTAATGTCAAGTCAACCGAAAGCCCTCTCTTACCTTCTGTAATCCACAGTCAAG 1175
    |||||
    514 AGAGCAAGTATTCTTCAACAACAGCAAGAAATGCTGCTTGAAGTGGTGGATTTG 573
    |||||
    1176 AGAGCAAGTATTCTTCAACAACAGCAAGAAATGCTGCTTGAAGTGGTGGATTTG 1235
    |||||
    574 AATGAGAAGACTTTGGGCTTATATCAGCTGTTTGGCATCCACCGGTGAGCATGAACAAT 633
    |||||
    1236 AATGAGAAGACTTTGGGCTTATATCAGCTGTTTGGCATCCACCGGTGAGCATGAACAAT 1295
    |||||
    634 AGCAACCACTTTGTGAATAGTACAAATATGGCTATGCGGCTTACCTTTTGAAGCAAGA 693
    |||||
    1296 AGCAACCACTTTGTGAATAGTACAAATATGGCTATGCGGCTTACCTTTTGAAGCAAGA 1355
    |||||
    694 AAGAACAAGGCTATCTGCGCTGAGACAGGACCGGTTTGTAGCAGGCTTTCCTTCAACAAT 753
    |||||
```

```

    |||||
    1356 AAGAACAAGGCTATCTGCTGTGACAGGACCGGTTGTGACAGGCTTTCCTTCAACAAT 1415
    |||||
    754 CTGCGAGACGCTGTACACCAACATTTCTTGCCCGCATTTGATGAACACAGAGGAGTCTGT 813
    |||||
    1416 CTGCGAGACGCTGTACACCAACATTTCTTGCCCGCATTTGTGTAAACACAGGAGTCTGT 1475
    |||||
    814 GACAAACAGACAGACACCTGTCCCAACGGTGGGCGCTAGCATGTGATGGCCAGCGACCT 873
    |||||
    1476 GAAACAGACAGACACCTGTCCCAACGGTGGGCGCTAGCATGTGATGGCCAGCGACCT 1535
    |||||
    874 GGACAAGACATGTTTGAAGACACACACATTTATAGACGATATCATATAGAAGGCAAG 933
    |||||
    1536 GGACAAGACATGTTTGAAGACACACACATTTATAGACGATATCATATAGAAGGCAAG 1595
    |||||
    934 GAGCTTATGCTCTGCTCTCCAGAGAGGTGACCGGCGCCAGTCTTGACCTCACACAGTGG 993
    |||||
    1596 GAGCTTATGCTCTGCTCTCCAGAGAGGTGACCGGCGCCAGTCTTGACCTCACACAGTGG 1655
    |||||
    994 GTGAACATGACAGATGTGAGCGTCCAGCTCAATGCCACACACAGTGAAGACGTGTAA 1053
    |||||
    1656 GTGAACATGACAGATGTGAGCGTCCAGCTCAATGCCACACAGTGAAGACGTGTAA 1715
    |||||
    1054 CCTGCCCTGGGCTACAGTGTGCGCGAGGCAATTTGATGAGTTCGGGCGCTCAATAT 1113
    |||||
    1716 CCTGCCCTGGGCTACAGTGTGCGCGAGGCAATTTGATGAGTTCGGGCGCTCAATAT 1775
    |||||
    1114 ACACAGGGAACACAGGAGGAGATCCATTTCTGGGACACTCTTGGGACACAGCTCTTGGA 1173
    |||||
    1776 ACACAGGGAACACAGGAGGAGATCCATTTCTGGGACACTCTTGGGACACAGCTCTTGGA 1835
    |||||
    1174 AAACATCTGAAGAGATTTAGAGTGTCAAAAACCAACCAATCTGCTTACAGTGA 1233
    |||||
    1836 AAACATCTGAAGAGATTTAGAGTGTCAAAAACCAACCAATCTGCTTACAGTGA 1895
    |||||
    1234 GAGCTACGATPACCAATCTTGGCAACAGATATTTGATGATTCAGATTTGTACCGGT 1293
    |||||
    1896 GAGCTACGATPACCAATCTTGGCAACAGATATTTGATGATTCAGATTTGTACCGGT 1955
    |||||
    1294 GGGTCTTGCCATACCTCTATCCCTGGGGAATTAACACACATGTGGGAGAGATTT 1353
    |||||
    1956 GGGTCTTGCCATACCTCTATCCCTGGGGAATTAACACACATGTGGGAGAGATTT 2015
    |||||
    1354 CGTAGGCAATTTAAAAACAATTTGACATTTATGGGATGAAGATATGACCTGTATATC 1413
    |||||
    2016 CGTAGGCAATTTAAAAACAATTTGACATTTATGGGATGAAGATATGACCTGTATATC 2075
    |||||
    1414 GCAGGTCTAAGCAATGTTATACACTTACATTTACACATTAAGAAGATACAGGCTGAG 1473
    |||||
    2076 GCAGGTCTAAGCAATGTTATACACTTACATTTACACATTAAGAAGATACAGGCTGAG 2135
    |||||
    1474 CGGTACGAGGACATCTACAATCTATGAGACACACACCTGTCTGATACATCCAACTC 1533
    |||||
    2136 CGGTACGAGGACATCTACAATCTATGAGACACACACCTGTCTGATACATCCAACTC 2195
    |||||
    1534 TTTCAGAGACCTTGTCTAAGCAATTTGCTACGACACAGTATGCCAATGAGAGTGTCC 1593
    |||||
    2196 TTTCAGAGACCTTGTCTAAGCAATTTGCTACGACACAGTATGCCAATGAGAGTGTCC 2255
    |||||
    1594 GAGCCTCCATCTCTCAAAAATCTAATAGCTTACCTTATCTTAATTCGGATGAGCA 1653
    |||||
    2256 GAGCCTCCATCTCTCAAAAATCTAATAGCTTACCTTATCTTAATTCGGATGAGCA 2315
    |||||
    1654 CCAATTTGCAAACTTTTGGGATGCTTGGACGCAAGCAAACTGAAATACAGAGTGGGA 1713
    |||||
    2316 CCAATTTGCAAACTTTTGGGATGCTTGGACGCAAGCAAACTGAAATACAGAGTGGGA 2375
    |||||
    1714 GAAAGTGTGAAGTATATTTGTAGGCGCTAAACCAAGAAATTCAGCAGAGACCAAGAC 1773
    |||||
    2376 GAAAGTGTGAAGTATATTTGTAGGCGCTAAACCAAGAAATTCAGCAGAGACCAAGAC 2435
    |||||
    1774 CATCAAACTTCTCAGTGTGAGAAATTCGAGGAGCTGTGTAGCTGAGCTGAGATTAATG 1833
    |||||
```

Db 2436 CATCAACCTTCCTCAGTGTGGAGAAATACGAGAGCTGTGACTGTGACGAGATTAATG 2495
OY 1834 TATAAGATGCTCTCGGAGAGAGGTTTATTTGGACAAAGGAAATAGTGGCTGAC 1893
Db 2436 TATAAGATGCTCTCGGAGAGAGGTTTATTTGGACAAAGGAAATAGTGGCTGAC 2555
OY 1894 AATGCAACATATATCTGATGATATTCAGATGCTGCTTACCGTGAATATACAGAAATAGA 1953
Db 2556 AATGCAACATATATCTGATGATATTCAGATGCTGCTTACCGTGAATATACAGAAATAGA 2615
OY 1954 TATTTTGGACACATGGAAGAGAGAACTTGAACCCGCTGATACACTAGCAATTTGAA 2013
Db 2616 TATTTTGGACACATGGAAGAGAGAACTTGAACCCGCTGATACACTAGCAATTTGAA 2675
OY 2014 GGAATTTCTCTCTTGTGAAGTGTCACTTGA 2049
Db 2676 GGAATTTCTCTCTTGTGAAGTGTCACTTGA 2711

RESULT 3
BC022604 1571 bp mRNA linear ROD 07-ANG-2002
LOCUS Mus musculus, Similar to N-acylsphingosine amidohydrolase 2, clone
DEFINITION IMAGE:4162513, mRNA, partial cds.
ACCESSION BC022604
VERSION BC022604.1 GI:18490446
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1571)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The J.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mdickson@stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
FEATURES
source
Clone distribution: MGC clone distribution information can be found
through the J.M.A.G.E. Consortium/LNL at: <http://image.jnl.gov>
Series: IRAX Plate: 43 Row: C Column: 22.
Location/Qualifiers
1. 1571
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="IMAGE:4162513"
/tissue_type="Liver, normal. 5 month old male mouse."
/clone_id="NCI CGAP_L19"
/lab_host="DH10B"
/note="vector: pCMV-SORT6"
1. 1493
/codon_start=3
/product="Similar to N-acylsphingosine amidohydrolase 2"
/protein_id="AAH22604.1"
/db_xref="GI:18490447"
/translation="KIVDNGEDLDLISFAIHPYSSNNHNFVSDNMGYAIYFEO
EKNGITLGGQGFVAGFASNSLGDVSPNLTGPHCVNTGSCNDSTCGNPGPSKMA
SGPGDMFESTHIIIGRIIYOKRELAYSASQVTPVLAHQVMWMTDVSVOJLNTHT
VTKRPAIGYSFAAGTIDVSGSLNTIGTTEBDPFWDLRLDILKPSSEIVECQPK

BASE COUNT 467 a 355 c 370 g 379 t
ORIGIN
PILHSGELTIPHPMDIVDQIVTVGSLAIATIPGELTTMSGRFREAIKREPALY
GKMDTVMVSSGNSYTHITTYEEEOARYSRASYIYGPHTLSAVIOJFRLAKAIA
TPPVANVSSGPEPPKKNLIALSLINADRIAGHFGVLOPAKREYGVVETIA
VGANRNSAENOTHOTELTVEYEDSVADMOJMYNDASBETFRYMHKGLISNATII
WHIPDAVPIGIRIRIFGNRRQELKRPVILAFBISSEFVVTT"
Query Match 72.9%; Score 1493; DB 10; Length 1571;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 557 TGAATGTGTGATTTGAATGGAAGAGACTTGGCTTATACAGTGGTTGCATCCACC 616
Db 1 TGAATGTGTGATTTGAATGGAAGAGACTTGGCTTATACAGTGGTTGCATCCACC 60
OY 617 CCGTGAGCATGAACAATATAGCAACACTTTGTGAATATAGTACATATAGGCTATGCGGCTT 676
Db 61 CCGTGAGCATGAACAATATAGCAACACTTTGTGAATATAGTACATATAGGCTATGCGGCTT 120
OY 677 ACCTTTGTGACAGAAAGAAACAAAGGCTATCTGCTGGACAGGACCGCTTGTAGCAG 736
Db 121 ACCTTTGTGACAGAAAGAAACAAAGGCTATCTGCTGGACAGGACCGCTTGTAGCAG 180
OY 737 GCTTTCCTTCATCAATATCTCGAGAGAGCTGACACCAATCTTGGCCCGCATTTGTGCA 796
Db 181 GCTTTCCTTCATCAATATCTCGAGAGAGCTGACACCAATCTTGGCCCGCATTTGTGCA 240
OY 797 ACACAGGGAGTCTTGTGACAAACAGACAGACAGACCTGTCCAAAGGCTGGCCTAGCATGT 856
Db 241 ACACAGGGAGTCTTGTGACAAACAGACAGACAGACCTGTCCAAAGGCTGGCCTAGCATGT 300
OY 857 GCATGGCCAGCGGAGCTGTGACAAAGCATTTTGAAGAGCACACAAATATAGAGGAGTCA 916
Db 301 GCATGGCCAGCGGAGCTGTGACAAAGCATTTTGAAGAGCACACAAATATAGAGGAGTCA 360
OY 917 TCTATCAGAAGGCCAAGAGAGCTGTATGCTTGCCTCCAGAGAGGTGACCGGCCAGTGC 976
Db 361 TCTATCAGAAGGCCAAGAGAGCTGTATGCTTGCCTCCAGAGAGGTGACCGGCCAGTGC 420
OY 977 TTGCAGCTCACCAGTGGTGAACATGATGATGAGCGTCCAGCTCAATGCCCACACACA 1036
Db 421 TTGCAGCTCACCAGTGGTGAACATGATGATGAGCGTCCAGCTCAATGCCCACACACA 480
OY 1037 CAGTGAAGACGTGTAAACCTGCGCCGCTGCTACAGTTTCCGACAGGACCAATGATGAG 1096
Db 481 CAGTGAAGACGTGTAAACCTGCGCCGCTGCTACAGTTTCCGACAGGACCAATGATGAG 540
OY 1097 TTTGGGCTCCTCAATATTTACACAGGAACTAGCGAAGGGAGATCCATTCTGGACACTTTC 1156
Db 541 TTTGGGCTCCTCAATATTTACACAGGAACTAGCGAAGGGAGATCCATTCTGGACACTTTC 600
OY 1157 GGGACAGCTCTTGGGAAAACCATCTGAAGAGATTGTAGAGTGCAGAAACCAACCA 1216
Db 601 GGGACAGCTCTTGGGAAAACCATCTGAAGAGATTGTAGAGTGCAGAAACCAACCA 660
OY 1217 TCGCTGTACAGTGGAGAGCGTACAGATACCAATCTTGGCAACAGATATTTGTGAG 1276
Db 661 TCGCTGTACAGTGGAGAGCGTACAGATACCAATCTTGGCAACAGATATTTGTGAG 720
OY 1277 TTCAGATTGTACCGTGTGGCTTGGCCATAGCTGTATCCCTGGGGAATTAACACCA 1336
Db 721 TTCAGATTGTACCGTGTGGCTTGGCCATAGCTGTATCCCTGGGGAATTAACACCA 780
OY 1337 TGTGGGACGAAGATTTGTGAGGCAATTAAGAAATTTGCATTTATGAGATGAGG 1396
Db 781 TGTGGGACGAAGATTTGTGAGGCAATTAAGAAATTTGCATTTATGAGATGAGG 840
OY 1397 ATATGACGCTGTATTCGAGAGCTTAGCAAGTATATACATTAACATTAACATTAAG 1456
Db 841 ATATGACGCTGTATTCGAGAGCTTAGCAAGTATATACATTAACATTAACATTAAG 900
OY 1457 AAGAAATACAGAGCTCAGCGTACGAGGCGAGATCTTACATCTATGAGACACACACCTGT 1516

Db		901	AAGAATACCGAGCTCAGGGGTACGAGGAGCATCTACAACTTATGGACACACACCCTGT	960
OY		1517	CTGCATATCATCCAACTCTTCAGAGACTTGCTAAGCAATTGCTACGAGACAGTAGCCA	1576
Db		961	CTGCATATCATCCAACTCTTCAGAGACCTGCTAAGGCAAAATTGCTACGAGACACATAGCCA	1020
OY		1577	ACAATGAGAGAGTGGGCCGAGGCCCTCAATTTCTCAAANAATCTAATAGCTTCACTTATTTCTTA	1636
Db		1021	ACAATGAGAGAGTGGGCCGAGGCCCTCAATTTCTCAAANAATCTAATAGCTTCACTTATTTCTTA	1080
OY		1637	ATATTGGCGATAGAGCACCAATTTGGCAAAACATTTTGGGAGTGTCTTCAGCCAGCAAAAAC	1696
Db		1081	ATATTGGCGATAGAGCACCAATTTGGCAAAACATTTTGGGAGTGTCTTCAGCCAGCAAAAAC	1140
OY		1697	CTGAATAACAGAGTGGGAGAACGTGGTGAAGTTATATTTGTAGGGCGCTAACCCAAAGAATT	1756
Db		1141	CTGAATAACAGAGTGGGAGAACGTGGTGAAGTTATATTTGTAGGGCGCTAACCCAAAGAATT	1200
OY		1757	CAGCAGAGAACCAACACCCTATCAACCTTCTCATCTGTGGAGAAATPACGAGACCTGTGAG	1816
Db		1201	CAGCAGAGAACCAACACCCTATCAACCTTCTCATCTGTGGAGAAATPACGAGACCTGTGAG	1260
OY		1817	CTGACTGGCAGATATATGTATPAACGATGGCTCTCTGGAGACGAGGTTTTATTTGGCACAAG	1876
Db		1261	CTGACTGGCAGATATATGTATPAACGATGGCTCTCTGGAGACGAGGTTTTATTTGGCACAAG	1320
OY		1877	GAATACHTGGGTCTGAGCAATGCAACAAATATACCTGGCATATTCGAGATACCTGCCCTACCTG	1936
Db		1321	GAATACHTGGGTCTGAGCAATGCAACAAATATACCTGGCATATTCGAGATACCTGCCCTACCTG	1380
OY		1937	GAACTCTACAGAAATTAAGATATTTTGGACACAATCGGAAGCAGAACTTCTGAAACCCGCTG	1996
Db		1381	GAACTCTACAGAAATTAAGATATTTTGGACACAATCGGAAGCAGAACTTCTGAAACCCGCTG	1440
OY		1997	TCATACAGACATTTGAAGAATTTCTCTCTCTTTGAAGTGGTACACTAG	2049
Db		1441	TCATACAGACATTTGAAGAATTTCTCTCTCTTTGAAGTGGTACACTAG	1493
RESULT 4				
ACL17816				
LOCUS	Mus musculus clone RP24-573C13, WORKING DRAFT SEQUENCE, 12			
DEFINITION	unordered pieces.			
ACCESSION	ACL17816			
VERSION	AC117816.2 GI:22267711			
KEYWORDS	HTG: HTGS PHASE1; HTGS_DRAFT.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryotes: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
TITLE	1 (bases 1 to 165407)			
JOURNAL	Birten,B., Nusbaum,C. and Lander,E.			
REFERENCE	Mus musculus, clone Rp24-573C13			
AUTHORS	Unpublished			
	2 (bases 1 to 165407)			
	Birten,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,			
	Anderson,S., Barne,N., Baselin,V., Bloom,T., Boguslavsky,L.,			
	Boukhalter,B., Brown,A., Camarota,J., Campopiano,A., Chang,J.,			
	Chazaro,B., Choepel,Y., Colangelo,W., Collins,S., Collamore,A.,			
	Cook,A., Coore,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,			
	Fato,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,			
	Glade,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,			
	Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,			
	Kamat,A., Karatsas,A., Kells,C., Labrecque,K., Lamazates,R.,			
	Landers,T., Lehoczkxy,J., Levine,R., Lindblad-Toh,K., Liu,G.,			
	Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,			
	McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menous,L.,			
	Milovov,T., Mlenda,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,			
	Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,			
	Oliver,J., Peterson,K., Phunkhang,P., Plerre,N., Poljata,V.,			
	Raymond,C., Retla,R., Rieback,H., Riley,R., Rise,C., Rogov,P.,			

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Roy, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schpack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, B., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 165407)
 Birren, B., Nisbam, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barua, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diz, J. S., Dodge, S., Fero, S., Ferreira, P., Fitzgeraid, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneses, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhkheng, P., Plerre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schpack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, B., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (16-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 16, 2002 this sequence version replaced g1:20128606.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 125854
 Center clone name: 573.C.13
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960721
 Consensus quality: 160616 bases at least Q40
 Consensus quality: 162867 bases at least Q20
 Consensus quality: 163827 bases at least Q20
 Insert size: 168000; agarose-fp
 Insert size: 164307; sum-of-contigs
 Quality coverage: 7.1 in Q20 bases; agarose-fp
 Quality coverage: 7.2 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 705: contig of 705 bp in length
 * 706 805: gap of 100 bp
 * 806 1789: contig of 984 bp in length
 * 1790 1889: gap of 100 bp
 * 1890 3396: contig of 1507 bp in length
 * 3397 3496: gap of 100 bp
 * 3497 4739: contig of 1243 bp in length
 * 4740 4839: gap of 100 bp
 * 4840 11073: contig of 6234 bp in length
 * 11074 11173: gap of 100 bp

Assembly program: Phrap; version 0.960731
 Consensus quality: 160616 bases at least Q40
 Consensus quality: 162967 bases at least Q30
 Consensus quality: 163827 bases at least Q20
 Insert size: 168000; agarose-fp
 Insert size: 164307; sum-of-coverage
 Quality coverage: 7.1 in Q20 bases; agarose-fp
 Quality coverage: 7.2 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 705: contig of 705 bp in length
 * 706 805: gap of 100 bp
 * 806 1789: contig of 984 bp in length
 * 1790 1889: gap of 100 bp
 * 1890 3396: contig of 1507 bp in length
 * 3397 3496: gap of 100 bp
 * 3497 4739: contig of 1243 bp in length
 * 4740 4839: gap of 100 bp
 * 4840 11073: contig of 6234 bp in length
 * 11074 11173: gap of 100 bp
 * 11174 18734: contig of 7561 bp in length
 * 18735 18834: gap of 100 bp
 * 18835 29191: contig of 10357 bp in length
 * 29192 29291: gap of 100 bp
 * 29292 42552: contig of 13261 bp in length
 * 42553 42652: gap of 100 bp
 * 42653 60018: contig of 17366 bp in length
 * 60019 60118: gap of 100 bp
 * 60119 81529: contig of 21411 bp in length
 * 81530 81629: gap of 100 bp
 * 81630 120500: contig of 38871 bp in length
 * 120501 120600: gap of 100 bp
 * 120601 165407: contig of 44807 bp in length.

FEATURES

source

1. 165407
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_id="RP24-573C13"
 /clone_lib="RPC1-24 Male Mouse BAC"
 1. 705
 misc_feature
 /note="assembly-fragment"
 806..1789
 misc_feature
 /note="assembly-fragment"
 1890..3396
 misc_feature
 /note="assembly-fragment"
 3497..4739
 misc_feature
 /note="assembly-fragment"
 4840..11073
 misc_feature
 /note="assembly-fragment"
 11174..18734
 misc_feature
 /note="assembly-fragment"
 18835..29191
 misc_feature
 /note="assembly-fragment"
 29292..42552
 misc_feature
 /note="assembly-fragment"
 42653..60018
 misc_feature
 /note="assembly-fragment"
 60119..81529
 misc_feature
 /note="assembly-fragment"
 81630..120500
 misc_feature
 /note="assembly-fragment"
 120601..165407
 misc_feature
 /note="assembly-fragment"
 165407..165407
 BASE COUNT 46319 a 32936 c 33090 g 51958 t 1104 others
 ORIGIN

Query Match 8.7%; Score 179; DB 2; Length 165407;
 Best Local Similarity 100.0%; Pred. No. 1.5e-95;
 Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

215 AGGCTCTGAAGACATAGAGATTAATATGCGCTCTGTATGAGAGACATATATCC 274
 Db 17534 AGGCTCTGAAGACATAGAGATTAATATGCGCTCTGTATGAGAGACATATATCC 17415
 QY 275 TGAGTCCATTCACACACACTCTGGCCACGAGGGTTTTCATATACACTATATAC 334
 Db 17474 TGAGTCCATTCACACACACTCTGGCCACGAGGGTTTTCATATACACTATATAC 17415
 QY 335 TCGCCACGAGGATTCACCAACCGACCTTTCAGTACATATCTCTGGATCATGAG 333
 Db 17414 TCGCCACGAGGATTCACCAACCGACCTTTCAGTACATATCTCTGGATCATGAG 17356

RESULT 6

AC113485

LOCUS

AC113485 184101 bp DNA linear HTG 06-JUN-2002

DEFINITION

Mus musculus clone RP23-346D12, WORKING DRAFT SEQUENCE, 17 ordered pieces.

ACCESSION

AC113485.3 GI:21327530

VERSION

HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

house mouse.

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE

Mus musculus, clone RP23-346D12

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 184101)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faru, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., McKernan, K., McEwen, P., Meneus, L., Milnova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Polara, V., Raymond, C., Retta, R., Riedack, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfay, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

3 (bases 1 to 184101)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faru, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwen, P., McKernan, K., Melidim, J., Meneus, L., Milnova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,

pieces.
AC113485 GI:21327530
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
SOURCE
house mouse.
ORGANISM
Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 184101)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, A., Allen, N., Brown, A., Camarata, J., Chang, J., Chazaro, B., Choquet, Y., Colangelo, M., Collins, S., Collumore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S., Fero, S., Ferrel, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menes, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thoman, N., Stojanovic, N., Struss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travs, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 184101)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, A., Allen, N., Anderson, S., Barne, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B., Choquet, Y., Colangelo, M., Collins, S., Collumore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S., Fero, S., Ferrel, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menes, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Stange-Thoman, N., Stojanovic, N., Struss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travs, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:21313862.
ALL repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L23776
Center clone name: 346.D.12
----- Summary Statistics -----
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 17831 bases at least Q40
Consensus quality: 181207 bases at least Q30
Consensus quality: 182003 bases at least Q20
Insert size: 183000; agarose-1p
Insert coverage: 182501; sum-of-contigs
Quality coverage: 6.7 in Q20 bases; agarose-1p
Quality coverage: 6.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 1186: contig of 1186 bp in length
1 1187 1286: gap of 100 bp
1 1287 2509: contig of 1223 bp in length
2510 2609: gap of 100 bp
2610 4200: contig of 1591 bp in length
4201 4300: gap of 100 bp
4301 6783: contig of 2483 bp in length
6784 6883: gap of 100 bp
6884 7958: contig of 1075 bp in length
7959 8058: gap of 100 bp
8059 10796: contig of 2738 bp in length
10797 10896: gap of 100 bp
10897 13878: contig of 2992 bp in length
13879 13978: gap of 100 bp
13979 15988: contig of 2010 bp in length
15989 16088: gap of 100 bp
16089 20675: contig of 4587 bp in length
20676 20775: gap of 100 bp
20776 25506: contig of 4731 bp in length
25507 25606: gap of 100 bp
25607 30961: contig of 5355 bp in length
30962 31061: gap of 100 bp
31062 41639: contig of 10578 bp in length
41640 41739: gap of 100 bp
41740 77584: contig of 35845 bp in length
77585 77684: gap of 100 bp
77685 107166: contig of 29482 bp in length
107167 107266: gap of 100 bp
107267 140582: contig of 33316 bp in length
140583 140682: gap of 100 bp
140683 174914: contig of 34232 bp in length
174915 175014: gap of 100 bp
175015 184101: contig of 9087 bp in length.
Location/Qualifiers
1. 184101
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-346D12"
/clone_1lb="RPci-23 Female Mouse BAC"
1. 1186
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
1287. 2509
/note="assembly_fragment"
2610. 4200
/note="assembly_fragment"
4301. 6783
/note="assembly_fragment"

```
misc_feature      684..7958
                  /note="assembly_fragment"
misc_feature      8059..10796
                  /note="assembly_fragment"
misc_feature      10897..13878
                  /note="assembly_fragment"
misc_feature      13979..15988
                  /note="assembly_fragment"
misc_feature      16089..20675
                  /note="assembly_fragment"
misc_feature      20776..22506
                  /note="assembly_fragment"
misc_feature      25607..30961
                  /note="assembly_fragment"
misc_feature      31062..41639
                  /note="assembly_fragment"
misc_feature      41740..77584
                  /note="assembly_fragment"
misc_feature      77685..107166
                  /note="assembly_fragment"
misc_feature      107267..140582
                  /note="assembly_fragment"
misc_feature      140683..174914
                  /note="assembly_fragment"
misc_feature      175015..184101
                  /note="assembly_fragment"
misc_feature      clone_end:T7
                  vector_side:right"
BASE COUNT      51100 a 39204 c 38438 g 53758 t 1601 others
ORIGIN
Query Match      3.5%; Score 72; DB 2; Length 184101;
Best Local Similarity 100.0%; Pred. No. 7.3e-31;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      147 AGATGGTCAATGCAATGTCATTGAGCGTGAACATATATGATTTCCCAACGACT 206
          |||||||
Db      30961 AGATGGTCAATGCAATGTCATTGAGCGTGAACATATATGATTTCCCAACGACT 30902
QY      207 GAGGTTGGAGCT 218
          |||||||
Db      30901 GAGGTTGGAGCT 30890

RESULT 8
E50433
LOCUS      E50433      2283 bp      DNA      linear      PAT 31-JAN-2002
DEFINITION      Novel protein.
ACCESSION      E50433
VERSION      E50433.1 GI:18633518
KEYWORDS      JP 2001057890-A/1.
SOURCE      Rattus norvegicus.
ORGANISM      Rattus norvegicus.
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                  Rattus.
REFERENCE      1 (bases 1 to 2283)
AUTHORS      Omori,A. and Ito,M.
JOURNAL      Novel protein.
PATENT      JP 2001057890-A 1 06-MAR-2001;
MITSUBISHI CHEM CORP
OS      Rattus norvegicus (rat)
PN      JP 2001057890-A/1
PD      06-MAR-2001
PR      23-AUG-1999 JP 1999235218

COMMENT
PI      AKIRA OMORI, MAKOTO ITO
PC      C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/80, PC
CC      C12N15/00, C12N5/00
FH      Key
FT      CDS      Location/Qualifiers
                  (1)..(2283).
```

```
source      1..2283
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
BASE COUNT      641 a 584 c 524 g 534 t
ORIGIN
Query Match      3.3%; Score 68; DB 6; Length 2283;
Best Local Similarity 100.0%; Pred. No. 1.8e-28;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1057 GCCCTGGGCTACAGTCTTCCGACAGCACAATGATGAGTTCGGCCCTCAATATTACA 1116
          |||||||
Db      1294 GCCCTGGGCTACAGTCTTCCGACAGCACAATGATGAGTTCGGCCCTCAATATTACA 1353
QY      1117 CAGGGAAC 1124
          |||||||
Db      1354 CAGGGAAC 1361

RESULT 9
AB057433
LOCUS      AB057433      2456 bp      mRNA      linear      ROD 12-JUL-2001
DEFINITION      Rattus norvegicus PAPKCD mRNA for ceramidase, complete cds.
ACCESSION      AB057433
VERSION      AB057433.1 GI:14701597
KEYWORDS
SOURCE      Rattus norvegicus kidney cDNA to mRNA.
ORGANISM      Rattus norvegicus
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                  Rattus.
REFERENCE      1
AUTHORS      Mitsuake,S., Tani,M., Okino,N., Mori,K., Ichinose,S., Omori,A.,
                  Iida,H., Nakamura,T. and Ito,M.
TITLE      Purification, characterization, molecular cloning, and subcellular
JOURNAL      J. Biol. Chem. 276 (28), 26249-26259 (2001)
MEDLINE      21365561
REFERENCE      2 (bases 1 to 2456)
AUTHORS      Ito,M.
JOURNAL      Direct Submission
TITLE      Submitted (14-MAR-2001) Makoto Ito, Kyushu University, Bioresource
          and Bioenvironmental Sciences; hakozaaki 6-10-1, higaashi-ku, Fukuoka
          812-8581, Japan (E-mail:makotoi@agr.kyushu-u.ac.jp,
          Tel:81-92-642-2900, Fax:81-92-642-2900)
FEATURES
            source
            1..2456
            Location/Qualifiers
            1..2456
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
            /tissue-type="kidney"
            1..2456
            /gene="PAPKCD"
            70..2355
            /gene="PAPKCD"
            /codon_start=1
            /product="ceramidase"
            /protein_id="PAB62033.1"
            /db_xref="GI:14701598"
            /translation="MAKRTFSLEAFLLIPLVMATAIVALLLLFVTSCTIENHKDS
            GNMWSTTGGPTTQSSPTQTPTTPDLPSPNSFGYVIGVARADCTQOVSDINM
            GYKNGCNOAGOLILRLFSRAFLIADPGSRMAFVSVELCMIISRLRLFLKRSKY
            GSVLRNDVILSATHTSGSPGFQVYLLIILASGSPNRTFOYVSGIVKSIDIAH
            LKRGKVLINRGANNOVINSRSPSYLONPSEARVSSPDKEVYVKKYLDINGEDG
            LISWFAVHPVSMNSNHLVNSDNGTAATILPEDEKNRGYLPQSGPFVAGRASNLGY
            SPNLGPHCVNTGESCDNRKSTCPGSGSPSCMASGPQDMESTHIIIGRIYVOKAEL
            HASASQEVTPGVLTIAHQVMNTDVSVOLNATHVKTCKALGYSFAAGTIDVSGNLI
            TGGTTEGNLFMDLRLDOLGKPSSEIIECQKPKPILITGEILPHEWOPDIYDIQIV
            TLGSLAIAALPGETTMSGRRLRAVAKKEPALGKMDNTVIAGLSWYTHYITTYE
            YKQARVAASTIYCPHTLSAYIOLFALAKAIAITDPYANNSGPEPEFKNLIGSLIP
            NIADRAPIGKQFGDVLDPAPKPEIRVGEVVEVYGANPKSAENQTHOTLTVEKID
            SVANWQIMHNDASWETREYVHKVGLISNATIHMHIPDTALPGVYIRIRYGHNNKQEL
            LKPAVILAFEGISPEEIVTT"
```

BASE COUNT

704 a 612 c 562 g 578 t

ORIGIN

Query Match 3.3%; Score 68; DB 10; Length 2456;
 Best Local Similarity 100.0%; Pred. No. 1.8e-28;
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1057 GCCCTGGGCTACAGTTTGGCCGAGCACAATTGATGAGTTGGGGCTCAATATTACA 1116
 |||||||
 DB 1363 GCCCTGGGCTACAGTTTGGCCGAGCACAATTGATGAGTTGGGGCTCAATATTACA 1422

OY 1117 CAGGGAC 1124
 |||||||

DB 1423 CAGGGAC 1430

RESULT 10
 E50435 539 bp DNA 1linear PAT 31-JAN-2002
 LOCUS E50435
 DEFINITION Novel protein.
 ACCESSION E50435
 VERSION E50435.1 GI:18633520
 KEYWORDS JP 2001057890-A/3.
 SOURCE synthetic construct.
 ORGANISM artificial sequences.
 REFERENCE 1 (bases 1 to 539)
 AUTHORS Omori, A. and Ito, M.
 TITLE Novel protein
 JOURNAL Patent: JP 2001057890-A 3 06-MAR-2001;
 MITSUBISHI CHEM CORP
 PN JP 2001057890-A/3
 OS Artificial Sequence
 PD 06-MAR-2001
 PF 23-AUG-1999 JP 1999235218
 PR
 PI AKIRA OMORI, MAKOTO ITO
 PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/80, PC
 C12N15/00, C12N5/00
 CC
 FH Key 1. Location/Qualifiers
 FT source 1.539
 FT Location/Qualifiers
 1.539
 /organism="synthetic construct"
 /db_xref="taxon:32630"

BASE COUNT 181 a 110 c 113 g 135 t

ORIGIN

Query Match 2.6%; Score 53; DB 6; Length 539;
 Best Local Similarity 100.0%; Pred. No. 2.1e-19;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1765 AACGAGACCATCAACCTCTCTCACTGTGAGAAATAGAGAGACTCTGAC 1817
 |||||||
 DB 1354 AACGAGACCATCAACCTCTCTCACTGTGAGAAATAGAGAGACTCTGAC 206

RESULT 11
 HUMD4C12M5 235 bp mRNA 1linear PRI 04-FEB-1999
 LOCUS HUMD4C12M5
 DEFINITION Human HepG2 partial cDNA, clone hmd4c12m5.
 ACCESSION D17062
 VERSION D17062.1 GI:598818
 KEYWORDS gene signature.
 SOURCE Homo sapiens Male cell_line:HepG2 cDNA to mRNA, clone_11b:Kiseru.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 235)
 AUTHORS Matoba, R., Okubo, K., Horii, N., Fukushima, A. and Matsubara, K.
 TITLE The addition of 5'-coding information to a 3'-directed cDNA library
 Improves analysis of gene expression

JOURNAL Gene 146 (2), 199-207 (1994)
 MEDLINE 94357437
 REFERENCE 2 (bases 1 to 235)
 AUTHORS Matoba, R.
 TITLE Direct Submission
 JOURNAL Submitted (21-JUL-1993) Ryo Matoba, Osaka University, Institute for
 Molecular and Cellular Bio; 1-3, Yamada-oka, Suita, Osaka 565,
 Japan (E-mail:matoba@inherit.imb.osaka-u.ac.jp,
 Tel:81-6-877-511(ex.3314), Fax:81-6-877-1922)
 Submitted (21-Jul-1993) to DDBJ by:
 Ryo Matoba
 COMMENT Molecular Microbiology and Genetics Lab.
 Research Institute of Innovative Technology for the Earth 9-2
 Kizugawadal Kizu-cyo,
 Soraku-gun, Kyoto
 Japan, 619-02
 Phone: 07747-5-2308
 Fax: 07747-5-2321.

FEATURES
 source 1.235
 Location/Qualifiers
 1.235
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /sex="Male"
 /cell_line="HepG2"
 /clone_11b="Kiseru"

BASE COUNT 64 a 46 c 57 g 68 t

ORIGIN

Query Match 2.3%; Score 47; DB 9; Length 235;
 Best Local Similarity 100.0%; Pred. No. 8.6e-16;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1933 CCTGAATCTACAGATAGATATTGGACACATCGAGACAGA 1979
 |||||||
 DB 102 CCTGAATCTACAGATAGATATTGGACACATCGAGACAGA 148

RESULT 12
 AF250847 2289 bp mRNA 1linear PRI 17-JUL-2000
 LOCUS AF250847
 DEFINITION Homo sapiens mitochondrial ceramidase mRNA, complete cds; nuclear
 gene for mitochondrial product.
 ACCESSION AF250847
 VERSION AF250847.1 GI:9246992
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 2289)
 AUTHORS El Bawab, S., Roddy, P., Qian, T., Bielawska, A., Lemasters, J.J. and
 Hannun, Y.A.
 TITLE Molecular cloning and characterization of a human mitochondrial
 ceramidase
 JOURNAL J. Biol. Chem. 275 (28), 21508-21513 (2000)
 MEDLINE 20347271
 PUBMED 10781606
 REFERENCE 2 (bases 1 to 2289)
 AUTHORS El Bawab, S. and Hannun, Y.A.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAR-2000) Biochemistry, Medical University of South
 Carolina, 114 Doughty Street, P.O. Box 250780, Charleston, SC
 29425, USA

FEATURES
 source 1.2289
 Location/Qualifiers
 1.2289
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1.2286
 /function="cleaves lipid ceramide into free fatty acid and
 sphingosine"
 /note="active in neutral to alkaline range"
 /codon_start=1
 /product="mitochondrial ceramidase"

```

/protein_id="AAFB6240.1"
/db_xref="GI:9246993"
/translation="MSAIVALLSLFTTGTIENHKLGGHFFSTTOSPPATOGSTA
AORSTATOHSTATQSTATQSPVLPPELSPFQNESSYHIGVRADCTGVADINLM
GYKSGONAGCILTRILYSRAFIAMEPDGSNRTVEVDIGAVSOLRLLEVLRLOSKY
GSLRRDNVLLSGTHHSRATFMAEPDGSNRTVEVDIGAVSOLRLLEVLRLOSKY
MKPKRIFLNKGNVGVQINRSPSYLQNPQSEARYPENIDKEMVLKAVDINGDLG
LISWFAIHVPMNNSNHLVNSDNGVYASYLEOEKRNKGLYEGGEPFAFASNGIDV
SPNLTGPCIWNTGSCDNANSTCPDGGSMCIAGPGQDMDSQIIGRAMYORAKEL
YASAOEYTGPIASAHQWMDVTWVNSHAKTCKPALCYSPACTIDGVCGLNT
TQGTGDEDPEDTTRDOLKPKSEIEKCHPKPRLILHTGSLKRPHPHPIVDQII
TLGSLALTAIPGEFTTMSGRRLREAVQAFASHMONTVVIGSCINVTYHTTYEE
YQAOEYEAATVIGPHALISAYIQLFRNLAKAIADVTYANLSGEPPEFKOLIVPLIP
SIDVRAPKRGFTGVDVLOPAKPEYVGEVAGNANVYONOTHOTFLVEKYEA
TSTSMOIVCNDASWETREYMHKGLGLSNATVEWHIPIPTAQIGYIRIRYFGHNKKODI
LKPAVILSFEETSAPFEVYTI"

BASE COUNT      664 a      553 c      515 g      557 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 8,9e-16;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1933 CCTGGAATCTACAGATAAGATATTGTCGACACATCGAAGCAGCA 1979
Db      2170 CCTGGAATCTACAGATAAGATATTGTCGACACATCGAAGCAGCA 2216

RESULT 13
AX207121      2507 bp      DNA      linear      PAT 30-AUG-2001
LOCUS      Sequence 1 from Patent WO0155410.
DEFINITION      AX207121
ACCESSION      AX207121
VERSION      AX207121.1 GI:15394925
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 2507)
AUTHORS      Hannu, Y.A. and el Bawab, S.
TITLE      Ceramund compositions and methods based thereon
JOURNAL      Patent: WO 0155410-A 1 02-AUG-2001;
MUSC Foundation For Research Development (US)

FEATURES
source
1. 2507
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
39..2324
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC60194.1"
/db_xref="GI:15394926"
/translation="MSAIVALLSLFTTGTIENHKLGGHFFSTTOSPPATOGSTA
AORSTATOHSTATQSTATQSPVLPPELSPFQNESSYHIGVRADCTGVADINLM
GYKSGONAGCILTRILYSRAFIAMEPDGSNRTVEVDIGAVSOLRLLEVLRLOSKY
GSLRRDNVLLSGTHHSRATFMAEPDGSNRTVEVDIGAVSOLRLLEVLRLOSKY
MKPKRIFLNKGNVGVQINRSPSYLQNPQSEARYPENIDKEMVLKAVDINGDLG
LISWFAIHVPMNNSNHLVNSDNGVYASYLEOEKRNKGLYEGGEPFAFASNGIDV
SPNLTGPCIWNTGSCDNANSTCPDGGSMCIAGPGQDMDSQIIGRAMYORAKEL
YASAOEYTGPIASAHQWMDVTWVNSHAKTCKPALCYSPACTIDGVCGLNT
TQGTGDEDPEDTTRDOLKPKSEIEKCHPKPRLILHTGSLKRPHPHPIVDQII
TLGSLALTAIPGEFTTMSGRRLREAVQAFASHMONTVVIGSCINVTYHTTYEE
YQAOEYEAATVIGPHALISAYIQLFRNLAKAIADVTYANLSGEPPEFKOLIVPLIP
SIDVRAPKRGFTGVDVLOPAKPEYVGEVAGNANVYONOTHOTFLVEKYEA
TSTSMOIVCNDASWETREYMHKGLGLSNATVEWHIPIPTAQIGYIRIRYFGHNKKODI
LKPAVILSFEETSAPFEVYTI"

BASE COUNT      720 a      582 c      567 g      638 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 8,9e-16;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      1933 CCTGGAATCTACAGATAAGATATTGTCGACACATCGAAGCAGCA 1979
Db      2208 CCTGGAATCTACAGATAAGATATTGTCGACACATCGAAGCAGCA 2254

RESULT 14
AL450382/c
LOCUS      AL450382
DEFINITION      Human DNA sequence from clone Rpl1-56A21 on chromosome 10, complete
sequence.
ACCESSION      AL450382
VERSION      AL450382.7 GI:21540029
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 162228)
AUTHORS      Peck, A.
TITLE      Direct Submission
JOURNAL      Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Jun 23, 2002 this sequence version replaced gi:1414881.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone configs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
Rpl1-56A21 is from the library RPl1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

FEATURES
source
1. 162228
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rpl1-56A21"
/clone_1lb="RPl1-11.1"
/codon_start=1
BASE COUNT      51458 a      32622 c      31866 g      46282 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 9,5e-16;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 62706 CCTGGAATCTACAGATAAGATATTTTGGACACATCGAGACGAGA 62660

RESULT 15
AL513306 176865 bp DNA linear HTG: 10-JUL-2001
LOCUS
DEFINITION Homo sapiens chromosome 10 clone RP11-532F4, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL513306
VERSION AL513306.10 GI:14530311
KEYWORDS HTG: HTGS_PHASE2; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 176865)
AUTHORS Howden, P.
TITLE Direct Submision
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jun 23, 2001 this sequence version replaced gi:13274890.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: bA532F4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 176570 bases at least Q40
Consensus quality: 176555 bases at least Q30
Consensus quality: 176694 bases at least Q20
Insert size: 176865; sum-of-ctrls
Insert size: 176572; 6.5% error; agarose-fp
Quality coverage: 9.23x in Q20 bases; sum-of-ctrls Quality
coverage: 9.25x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. 176865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-532F4"
/clone_1lb="RP11-11.2"
1. 176865
/note="assembly fragment: 04692"
BASE COUNT 53967 a 31937 c 33758 g 57203 t
ORIGIN

Query Match 2.3%; Score 47; DB 2; Length 176865;
Best Local Similarity 100.0%; Pred. No. 9.5e-16;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1933 CCTGGAATCTACAGATAAGATATTTTGGACACATCGAGACGAGA 1979
DB 32904 CCTGGAATCTACAGATAAGATATTTTGGACACATCGAGACGAGA 32950

RESULT 16
AL589794 178899 bp DNA linear PRI 18-SEP-2001
LOCUS
DEFINITION Human DNA sequence from clone RP11-564C4 on chromosome 10, complete
sequence.
ACCESSION AL589794
VERSION AL589794.7 GI:14330084

KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 178899)
AUTHORS Peck, A.
TITLE Direct Submision
JOURNAL Submitted (18-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jun 8, 2001 this sequence version replaced gi:13992363.
COMMENT During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
<http://www.sanger.ac.uk/Projects/Celegans/wormpep> This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>
RP11-564C4 is from the library RP11-2 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-564C4 The true
left end of clone RP11-532F4 is at 81721 in this sequence.
Location/Qualifiers
1. 178899
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-564C4"
/clone_1lb="RP11-11.2"
BASE COUNT 51555 a 36501 c 36785 g 54058 t
ORIGIN

Query Match 2.3%; Score 47; DB 9; Length 178899;
Best Local Similarity 100.0%; Pred. No. 9.5e-16;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1933 CCTGGAATCTACAGATAAGATATTTTGGACACATCGAGACGAGA 1979
DB 114624 CCTGGAATCTACAGATAAGATATTTTGGACACATCGAGACGAGA 114670

RESULT 17
AC055726_2
WPCOMMENT
Sequence split into 5 fragments
Fragment Name Begin End
AC055726_0 1 110000
AC055726_1 100001 210000
AC055726_2 200001 310000
AC055726_3 300001 410000
AC055726_4 400001 456720
Continuation (3 of 5) of AC055726 from base 200001 (AC055726 Homo sapiens chromosome
10, complete sequence).
Query Match 1.9%; Score 39; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 6.5e-11;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1933 CCTGCAATCTACAGATATATTTGGACACATCGG 1971
 Db 96512 CCTGCAATCTACAGATATATTTGGACACATCGG 96550

COMMENT

of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:18092937.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: CNPR

Center clone name: CH230-142G16

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 57464 bases at least Q40

Consensus quality: 61201 bases at least Q30

Consensus quality: 64573 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drafter_data.html).

* NOTE: This is a "working draft" sequence. It currently
 consists of 48 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

```

1      1226: contig of 1226 bp in length
*      1227      1326: gap of unknown length
*      1327      2590: contig of 1264 bp in length
*      2591      2690: gap of unknown length
*      2691      3839: contig of 1149 bp in length
*      3840      3939: gap of unknown length
*      3940      5328: contig of 1389 bp in length
*      5329      5429: gap of unknown length
*      5429      6596: contig of 1168 bp in length
*      6597      6696: gap of unknown length
*      6697      7891: contig of 1195 bp in length
*      7892      7992: gap of unknown length
*      7992      9247: contig of 1255 bp in length
*      9247      9347: gap of unknown length
*      9347      10448: contig of 1102 bp in length
*      10449      10549: gap of unknown length
*      10549      11837: contig of 1289 bp in length
*      11838      11937: gap of unknown length
*      11937      13599: contig of 1662 bp in length
*      13600      13699: gap of unknown length
*      13700      15116: contig of 1417 bp in length
*      15117      15216: gap of unknown length
*      15217      16273: contig of 1057 bp in length
*      16274      16374: gap of unknown length
*      16374      17450: contig of 1077 bp in length
*      17451      17550: gap of unknown length
*      17551      18836: contig of 1286 bp in length
*      18837      20118: gap of unknown length
*      20119      20218: contig of 1182 bp in length
*      20219      21463: contig of 1245 bp in length
*      21464      22745: gap of unknown length
*      22745      22745: contig of 1182 bp in length
*      22746      22846: gap of unknown length
*      22846      23877: contig of 1032 bp in length
*      23878      23978: gap of unknown length
*      23978      25337: contig of 1360 bp in length
*      25338      26655: gap of unknown length
*      26655      26755: contig of 1218 bp in length
*      26756      26756: gap of unknown length
*      26756      28546: contig of 1791 bp in length
*      28547      28646: gap of unknown length
*      28647      30123: contig of 1477 bp in length
*      30124      30223: gap of unknown length

```

RESULT 18
 AC105714
 LOCUS AC105714
 DEFINITION Rattus norvegicus clone CH230-142G16, *** SEQUENCING IN PROGRESS
 AC105714
 AC105714.2 GI:21736609
 HTG: HTGS.PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 105838)
 Murny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayalew,M., Banks,T.,
 Barbata,J., Benton,J., Blinag,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowe,S., Bileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buiay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,F., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratoch,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichteberg,O., Lieu,C., Liu,J., Liu,W., Louiseded,H.,
 Loado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okuwonu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A.A., Syatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.
 Unpublished
 Direct Submission
 2 (bases 1 to 105838)
 Worley,K.C.
 Direct Submission
 Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 105838)
 Worley,K.C.
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department

30224 32496: contig of 2273 bp in length
 32497 32596: gap of unknown length
 32597 34038: contig of 1442 bp in length
 34039 34138: gap of unknown length
 34139 36429: contig of 2291 bp in length
 36430 36529: gap of unknown length
 36530 38289: contig of 1770 bp in length
 38300 38399: gap of unknown length
 38400 40786: contig of 2387 bp in length
 40887 42807: gap of unknown length
 42808 42807: contig of 1921 bp in length
 42808 42807: gap of unknown length
 42808 45554: contig of 2647 bp in length
 45555 45654: gap of unknown length
 45655 47283: contig of 1628 bp in length
 47283 47382: gap of unknown length
 47383 49177: contig of 1795 bp in length
 49178 49277: gap of unknown length
 49278 51292: contig of 2015 bp in length
 51293 51392: gap of unknown length
 51393 52978: contig of 1586 bp in length
 52979 53078: gap of unknown length
 53079 55113: contig of 2035 bp in length
 55114 55213: gap of unknown length
 55214 56557: contig of 1344 bp in length
 56558 56657: gap of unknown length
 56659 59123: contig of 2466 bp in length
 59124 59223: gap of unknown length
 59224 61914: contig of 2691 bp in length
 61915 62014: gap of unknown length
 62015 65172: contig of 3158 bp in length
 65173 65272: gap of unknown length
 65273 67490: contig of 2218 bp in length
 67491 67590: gap of unknown length
 67591 71140: contig of 3550 bp in length
 71141 71240: gap of unknown length
 71241 73628: contig of 2388 bp in length
 73629 73728: gap of unknown length
 73729 76582: contig of 2854 bp in length
 76583 76682: gap of unknown length
 76683 79308: contig of 2626 bp in length
 79309 79408: gap of unknown length
 79409 84267: gap of unknown length
 84268 84366: gap of unknown length
 84367 89469: contig of 5103 bp in length
 89470 89569: gap of unknown length
 89570 94790: contig of 5221 bp in length
 94791 94890: gap of unknown length
 94891 99546: contig of 4656 bp in length
 99547 99646: gap of unknown length
 99647 105838: contig of 6192 bp in length.

FEATURES
 source 1.105838
 Location/Qualifiers
 1.105838
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-142G16"

BASE COUNT 28706 a 19526 c 20397 g 28661 t 8548 others

ORIGIN

Query Match 1.8%: Score 36; DB 2; Length 105838;
 Best Local Similarity 100.0%; Pred. No. 4.2e-09;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1623 TTCACCTATTCCTAATATTCGCGATAGACACCAAT 1658
 ||||||||||||||||||||||||||||||||||||||||
 Db 69354 TTCACCTATTCCTAATATTCGCGATAGACACCAAT 69389

RESULT 19 105838 bp DNA linear HTG 13-JUL-2002
 AC105714 C
 LOCUS Rattus norvegicus clone CH230-142G16, *** SEQUENCING IN PROGRESS
 DEFINITION *** 48 unordered pieces.

ACCESSION AC105714
 VERSION AC105714.2 GI:21736609
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 (bases 1 to 105838)
 Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C.,
 Alshrocks,S.L., Amaralungu,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbrale,J., Benton,J., Bimaye,K., Blankenburg,K., Bonin,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhaq,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dedetich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.D., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homel,F., Howard,S., Huber,J., Huijk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolyvet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratochvil,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,T., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Louised,H.,
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Matinez,E.,
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabab,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokoko,S., Ogun,M., Okunolu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

Direct Submission
 Unpublished
 2 (bases 1 to 105838)
 Worley,K.C.
 Direct Submission
 Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 105838)
 Worley,K.C.
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:18092937.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 Project information
 Center project name: GNR
 Center clone name: CH230-142G16
 ----- Summary Statistics

Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 57464 bases at least Q40
 Consensus quality: 61201 bases at least Q30
 Consensus quality: 64573 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see <http://www.hgsc.bcm.tmc.edu/docs/genbank.draft.data.html>).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 48 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

```

1      1226: contig of 1226 bp in length
*      1227      1326: gap of unknown length
*      1327      2590: contig of 1264 bp in length
*      2591      2690: gap of unknown length
*      2691      3839: contig of 1149 bp in length
*      3840      3939: gap of unknown length
*      3940      5328: contig of 1389 bp in length
*      5329      5429: gap of unknown length
*      5429      6596: contig of 1168 bp in length
*      6597      6696: gap of unknown length
*      6697      7891: contig of 1195 bp in length
*      7892      7991: gap of unknown length
*      7992      9246: contig of 1255 bp in length
*      9247      10448: contig of 1102 bp in length
*      10449      11837: gap of unknown length
*      11838      11937: contig of 1289 bp in length
*      11938      13599: contig of 1662 bp in length
*      13600      13699: gap of unknown length
*      13700      15116: contig of 1417 bp in length
*      15117      15216: gap of unknown length
*      15217      16273: contig of 1057 bp in length
*      16274      16374: gap of unknown length
*      16374      17450: contig of 1077 bp in length
*      17451      17550: gap of unknown length
*      17551      18836: contig of 1286 bp in length
*      18837      18936: gap of unknown length
*      18937      20119: contig of 1182 bp in length
*      20119      20218: gap of unknown length
*      20219      21463: contig of 1245 bp in length
*      21464      21563: gap of unknown length
*      21564      22745: contig of 1182 bp in length
*      22746      22845: gap of unknown length
*      22846      23877: contig of 1032 bp in length
*      23878      23977: gap of unknown length
*      23978      25337: contig of 1360 bp in length
*      25338      25437: gap of unknown length
*      25438      26655: contig of 1218 bp in length
*      26656      26755: gap of unknown length
*      26756      28546: contig of 1791 bp in length
*      28547      28646: gap of unknown length
*      28647      30123: contig of 1477 bp in length
*      30124      30223: gap of unknown length
*      30224      32496: contig of 2273 bp in length
*      32497      32596: gap of unknown length
*      32597      34038: contig of 1442 bp in length
*      34039      34138: gap of unknown length
*      34139      36429: contig of 2291 bp in length
*      36430      36529: gap of unknown length
*      36530      38299: contig of 1770 bp in length
*      38300      38399: gap of unknown length
*      38400      40786: contig of 2387 bp in length
*      40787      40886: gap of unknown length
*      40887      42807: contig of 1921 bp in length
*      42808      42907: gap of unknown length

```

```

*      42908      45554: contig of 2647 bp in length
*      45555      45654: gap of unknown length
*      45655      47282: contig of 1628 bp in length
*      47283      47382: gap of unknown length
*      47383      49177: contig of 1795 bp in length
*      49178      49277: gap of unknown length
*      49278      51292: contig of 2015 bp in length
*      51293      51392: gap of unknown length
*      51393      52978: contig of 1586 bp in length
*      52979      53078: gap of unknown length
*      53079      53113: contig of 2035 bp in length
*      53114      55213: gap of unknown length
*      55214      56557: contig of 1344 bp in length
*      56558      56657: gap of unknown length
*      56658      59123: contig of 2466 bp in length
*      59124      59223: gap of unknown length
*      59224      61914: contig of 2691 bp in length
*      61915      62014: gap of unknown length
*      62015      65172: contig of 3158 bp in length
*      65173      65272: gap of unknown length
*      65273      67490: contig of 2218 bp in length
*      67491      67591: gap of unknown length
*      67591      71140: contig of 3350 bp in length
*      71141      71240: gap of unknown length
*      71241      73628: contig of 2388 bp in length
*      73629      73728: gap of unknown length
*      73729      76582: contig of 2854 bp in length
*      76583      76682: gap of unknown length
*      76683      79308: contig of 2626 bp in length
*      79309      84266: gap of unknown length
*      84267      84365: gap of unknown length
*      84367      89469: contig of 5103 bp in length
*      89470      89569: gap of unknown length
*      89570      94790: contig of 5221 bp in length
*      94791      94890: gap of unknown length
*      94891      99546: contig of 4656 bp in length
*      99547      99646: gap of unknown length
*      99647      105838: contig of 6192 bp in length.

```

FEATURES
 source
 1..105838
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-142G16"

BASE COUNT
 28706 a 19526 c 20397 g 28661 t 8548 others

Query Match
 Best Local Similarity 100.0%; Pred. No. 1.1e-06;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1732 TTGTAGGCGCTAACCCAAAGATTTCAGCAGA 1763
 Db 52897 TTTGTAGGCGCTAACCCAAAGATTTCAGCAGA 52866

RESULT 20
 AL583858 187130 bp DNA linear HTG 10-JUL-2001
 AL583858/C
 LOCUS
 DEFINITION Homo sapiens chromosome 10 clone RP11-78J11, *** SEQUENCING IN
 PROGRESS ***, in ordered pieces.
 ACCESSION AL583858.9 GI:14575341
 VERSION
 KEYWORDS HTG; HTGS_PHASE2; HTGS_CANCELLED.
 SOURCE
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 1 (bases 1 to 187130)
 BROWN, J.
 Direct Submission
 Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone .

COMMENT requests: clonerequest@sanger.ac.uk
On Jun 28, 2001 this sequence version replaced gi:14330076.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: bA78J11
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 187080 bases at least Q40
Consensus quality: 187130 bases at least Q30
Insert size: 187130; sum-of-ctrligs
Insert size: 175296; 20.9% error; agarose-1p
Quality coverage: 7.08x in Q20 bases; sum-of-ctrligs Quality
coverage: 7.69x in Q20 bases; agarose-1p

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. 187130
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_11b="RP11-78J11"
/clone_11b="RP11-11.1"
1. 187130
/note="assembly-fragment:03017"
misc-feature 62248 a 37239 c 35291 g 53352 t
BASE COUNT
ORIGIN

Query Match 1.5%; Score 30; DB 2; Length 187130;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1682 TGCAGCCAGCAAAACCTGAATACAGTGG 1711
4136 TGCAGCCAGCAAAACCTGAATACAGTGG 4107

RESULT 21
AX207126 38 bp DNA linear PAT 30-AUG-2001
LOCUS Sequence 6 from Patent WO0155410.
DEFINITION AX207126
ACCESSION AX207126
VERSION AX207126.1 GI:15394932
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 38)
AUTHORS Hannun, Y.A. and el Bawab, S.
TITLE Ceramide compositions and methods based thereon
JOURNAL Patent: WO 0155410-A 6 02-AUG-2001;
MUSC Foundation For Research Development (US)
Location/Qualifiers
1. 38
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="primer"
BASE COUNT 11 a 10 c 10 g 7 t
ORIGIN

Query Match 1.2%; Score 25; DB 6; Length 38;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1682 TGCAGCCAGCAAAACCTGAATACAG 1706
Db 14 TGCAGCCAGCAAAACCTGAATACAG 38

RESULT 22
OSJNB00174 164805 bp DNA linear PLN 17-JUL-2002
LOCUS Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBa0058K23,
DEFINITION complete sequence.
ACCESSION AL662970
VERSION AL662970.2 GI:21741993
KEYWORDS HTG.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriarthroideae; Oryzaceae; Oryza.
REFERENCE 1
AUTHORS Han, B., Feng, Q., Huang, Y.C., Li, Y., Zhu, J.J., Zhao, Q., Hu, X.,
Liu, Y.L., Mu, J., Yu, Z., Chen, L., Fan, D.L., Meng, Q.D., Zhang, L.,
Lu, Y.Q., Yu, S.L., Liu, X.H., Lu, T.T., Zhang, Y.J., Lu, Y., Li, C.,
Li, T., Zhang, Y., Hu, H., Jia, P.X., Qian, Y.M., Ying, K., Zhou, B.,
Chen, Z.H., Hao, P., Zhang, L., Wu, M., Zhang, R.Q., Guan, J.P., Fu, G.,
Wang, S.T., Ren, S.X., Lv, G., Lin, W., Gu, W.Q., Zhu, G.F., Tu, Y.F.,
Jia, J., Yin, H.F., Zhang, Y., Cai, Z., Chen, J., Kang, H., Chen, X.Y.,
Shao, C.Y., Sun, Y., Hu, Q.P., Zhang, X.L., Zhang, W., Wang, L.J.,
Ding, C.W., Sheng, H.H., Gu, J.L., Chen, S.T., Ni, L., Zhu, F.H. and
Hong, G.F.
Direct Submission
Submitted (27-JUN-2002) Han Bin, National Center for Gene Research,
Chinese Academy of Sciences, 500# Cao Bao Road, Shanghai 200233,
CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
bhan@ncgr.ac.cn
Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
clone: OSJNBa0058K23.
On Jul 12, 2002 this sequence version replaced gi:17998482.
Web site: http://www.ncgr.ac.cn
----- Summary Statistics
Assembly program: phrap

NOTE: This is a complete sequence.
Genes were identified by a combination of several methods: Gene
prediction programs including Fgenesh (<http://www.softberry.com/>),
genscan (<http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHM
(<http://genemark.biology.gatech.edu/genemark/>), tRNAscan-SE (Sean
Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>), searches of the
complete sequence against NCBI non-redundant protein database (nr)
(<ftp://ncbi.nlm.nih.gov/blast/db>) and the EST database at NCGR.
Location/Qualifiers
1. 164805
/organism="Oryza sativa"
/variety="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:4530"
/chromosome="4"
/clone="OSJNBa0058K23"
/clone_11b="CUG1-OSJNBa"
BASE COUNT 47435 a 34481 c 35046 g 47843 t
ORIGIN

Query Match 1.1%; Score 23; DB 8; Length 164805;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1676 ATGCTCTGCAGCCAGCAAAACCT 1698
Db 160127 ATGCTCTGCAGCCAGCAAAACCT 160149

RESULT 23
AC012131 175330 bp DNA linear HTG 14-SEP-2000
LOCUS

```
DEFINITION Homo sapiens clone RP11-5E11, WORKING DRAFT SEQUENCE, 4 unordered
pieces.
ACCESSION AC012131
VERSION AC012131.5 GI:10122074
KEYWORDS HTG, HTGS_PHASE1, HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 175330)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens, clone RP11-5E11
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175330)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckery, R., Boguslavsky, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
Cooke, P., Dearlano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karates, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Neldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirelli, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 14, 2000 this sequence version replaced gi:18576235.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2849
Center clone name: 5_E_11
----- Summary Statistics
Sequencing vector: M13, M77815, 100% of reads
Sequencing vector: Plasmid, n/a; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 172203 bases at least Q40
Consensus quality: 173820 bases at least Q30
Consensus quality: 174439 bases at least Q20
Insert size: 157000; agarose-IP
Insert size: 175030; sum-of-contents
Quality coverage: 7.4 in Q20 bases; agarose-IP
Quality coverage: 6.6 in Q20 bas.
NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
777769: contig of 77769 bp in length
77770 77869: gap of 100 bp
77870 79091: contig of 1222 bp in length
79092 79191: gap of 100 bp
79192 139104: contig of 59913 bp in length
139105 139204: gap of 100 bp
139205 175330: contig of 36126 bp in length.
Location/Qualifiers
1..175330
/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone="RP11-5E11"
/clone_lib="RPCT-11 Human Male BAC"
1..77769
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
77870..79091
/note="assembly_fragment"
79192..139104
/note="assembly_fragment"
139205..175330
/note="assembly_fragment"
clone_end:T7
vector_side:right"

BASE COUNT 52740 a 33512 c 34651 g 54127 t 300 others
ORIGIN
Query Match 1.1%; Score 23; DB 2; Length 175330;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

940 TATGCTCTGCTCCACGAGGT 962
|||||
Db 169446 TATGCTCTGCTCCACGAGGT 169468

RESULT 24
FI93462S02 168 bp DNA linear PRI 12-NOV-2000
LOCUS Pan troglodytes dipeptidyl carboxy peptidase 1 (DCP1) gene, exon 2.
DEFINITION AF193463
ACCESSION AF193463.1 GI:11138388
VERSION
KEYWORDS
SEGMENTS
2 of 26
SOURCE Pan troglodytes.
ORGANISM Pan troglodytes.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1 (bases 1 to 168)
AUTHORS Dufour, C., Casane, D., Denton, D., Wickings, J., Corvol, P. and
Jeunemaitre, X.
Human-chimpanzee DNA sequence variation in the four major genes of
the renin angiotensin system
Genomics 69 (1), 14-26 (2000)
11013071
MEDLINE 20469400
PUBMED 2 (bases 1 to 168)
REFERENCE Dufour, C., Corvol, P. and Jeunemaitre, X.
AUTHORS Submitted (12-OCT-1999) INSERM U36, College de France, 3 Rue d'Ulm,
Paris 75005, France
JOURNAL Location/Qualifiers
1..168
/organism="Pan troglodytes"
/db_xref="taxon:9598".
1..168
/gene="DCP1"
/number=2
BASE COUNT 34 a 51 c 58 g 25 t
ORIGIN
Query Match 1.1%; Score 22; DB 9; Length 168;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

922 CAGAGGCCAAGAGCTGATG 943
|||||
Db 46 CAGAGGCCAAGAGCTGATG 67

RESULT 25
AR137383
```

LOCUS AR137383 4020 bp DNA linear PAT 16-JUN-2001
DEFINITION Sequence 130 from patent US 6197505.
ACCESSION AR137383
KEYWORDS AR137383.1 GI:14478892
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4020)
AUTHORS Norberg/L.Torjorn., Andersson,M.Kristina. and
Lindstrom,P.Harry,Rutger.
TITLE Methods for assessing cardiovascular status and compositions for
use thereof
JOURNAL Patent: US 6197505-A 130 06-MAR-2001;
FEATURES
source Location/Qualifiers
BASE COUNT 857 a 1261 c 1174 g 728 t
ORIGIN
Query Match 1.1%; Score 22; DB 6; Length 4020;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 922 CAGAGGCCAGAGAGCTGTATG 943
Db 317 CAGAGGCCAGAGAGCTGTATG 338
RESULT 26
LOCUS HUMA1CEB 4020 bp mRNA linear PRI 30-OCT-1994
DEFINITION Human angiotensin I-converting enzyme mRNA, complete cds.
ACCESSION J04144
VERSION J04144.1 GI:178285
KEYWORDS angiotensin converting enzyme; dipeptidyl carboxypeptidase.
SOURCE Human endothelial cell, cDNA to mRNA, clones lambda-HEC1922,
lambda-HEC2111, and lambda-CHD732.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 4020)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE Soubrier,F., Albenc-Gelas,F., Hubert,C., Allegri,J., John,M.,
Tregear,G. and Corvol,P.
TWO putative active centers in human angiotensin I-converting
enzyme revealed by molecular cloning
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 85 (24), 9386-9390 (1988)
MEDLINE 89071703
PUBMED 2849100
COMMENT Draft entry and computer-readable sequence [1] kindly submitted by
F.Soubrier 04-JAN-1989.
FEATURES
source Location/Qualifiers
BASE COUNT 1. .4020
ORIGIN /organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17q23"
1. .4020
/gene="DCP1"
1. .109
/note="DCP1"
/note="angiotensin I-converting enzyme signal peptide"
23. .3943
/gene="DCP1"
/note="angiotensin I-converting enzyme precursor (EC
3.4.15.1)"
3.4.15.1"
/codon_start=1
/protein_id="AA51684.1"
/db_xref="GI:178286"
/db_xref="GDB:600-119-840"
/translation="MGAAGRGPGCLLPPLPLLLPPOPALPDGPGNFSADAE
GAQLFAQSYNSAEQVLFQSVASNAHNTITAENARROEALISOEAAEAGKAK
ELYEPIMQNTDPPQRTIIGAVRTIGSANLPLAKROQYNALLSNRSRYSTKVCILN
KTAICWSLDPDLTNILASSRSYAMLLFMEGHNAAGLPKLFYEDFTALSNEAYKOD

GFTTGAVNRSYNSPTFEEDLEHLVQOLEPLYLNLHAFVRALHRRGDRYINLRGP
IPAHLLGDMMAQSWENTYDMVVPFDPKMLDVTSMLOQGNATMFRVAEEFTSLE
LSPMPBEWEGSMLEKPADREVVCASAMDFNKRDFIKOCTVTMDQLSTVHEH
GHIOYLOYKQDLPVSLRGANGFHEALIGDVALSVSPENHKLIGLDROYICPVT
INVILKMALEKTAFLPFGYLDQMRGVRGVSSEGTSPSRVMDWMTYRTKOGICPVT
NETHFDAGAKKEHVPNVTPYIRFYVSFVLOFQFHEALCKRAGEGLEHOCDIYRSKAG
AKRKVLQAGSSRPQOEVLKMDVADLDAOLPLKTFQVOTMLOEONQNEVLGMP
EYQMPLEPDNPESIDLVTDENAEASKVEEYDRTSQVWNEAYEAMNNTYET
SKILLQNMQIANHTLKYGTOARKFDVNOLOMTTIRIKKVOYDLERAPALPOELE
NILLDMETTYSVATYCHNGSCLOEPDITVMATSRKYEEDLLAMEGRMRKAGRAI
LOEYRYVELLINQAAKLNGYVADGSMRSRETPSEDOLELFOELOPLYLNLHAY
RALHRYVGAOIHINEGPYPAHLINMAQTSNTYDLYVPPSA.PSMDTEAMTKOG
WTPRRMFKRADPFTSLGLPVPPEPWRKSMLEKRTDGEVYCHASAMDFYNGKPRI
KOCITVNLLEDLVVAHEMGIQYEMQYKDLVPALREGANPGEHEAIGDVALSVTPK
HLSLNLSSSEGSDEHDINFLMKALDKIATIPSYLDQMRGVRGVSSEGTSPSRVMDWMTYRTKOGICPVT
EWMSLRLKQGLCPVPPTGDFDPGAKRIPSSVPIRYEVSFTIOPFHALCOAA
GHTGPLKCDIYOSKEAGORLATAMKLSRFPAMOLITOPNMSASMYSTKPL
LDMRTRENLEHGEKIGMPQYNTSPNSARESGPLPSGSEFLGLDLDAQARVQGMWL
LFLGIALIVATIGLSORLEFSTRSLRHRSRHPORGEVELRHS"
110. .3940
mat_peptide
/gene="DCP1"
/product="angiotensin I-converting enzyme"
BASE COUNT 857 a 1261 c 1174 g 728 t
ORIGIN
Query Match 1.1%; Score 22; DB 9; Length 4020;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 922 CAGAGGCCAGAGAGCTGTATG 943
Db 317 CAGAGGCCAGAGAGCTGTATG 338
RESULT 27
LOCUS AX429555 4022 bp DNA linear PAT 21-JUN-2002
DEFINITION Sequence 18 from Patent WO0226820.
ACCESSION AX429555
VERSION AX429555.1 GI:21540823
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE Cohen,D., Dengler,U.J., Finelli,A.L., Freuler,F., Konsolaki,M.,
Reinhardt,M.W. and Zusman,S.
Transgenic drosophila melanogaster expressing beta amyloid
JOURNAL Patent: WO 0226820-A 18 04-APR-2002;
NOVARTIS ERFIND VERWALT GMBH (AT)
FEATURES
source Location/Qualifiers
BASE COUNT 1. .4022
ORIGIN /organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17q23"
1. .4020
/gene="DCP1"
1. .109
/note="DCP1"
/note="angiotensin I-converting enzyme precursor (EC
3.4.15.1)"
3.4.15.1"
/codon_start=1
/protein_id="AA51684.1"
/db_xref="GI:178286"
/db_xref="GDB:600-119-840"
/translation="MGAAGRGPGCLLPPLPLLLPPOPALPDGPGNFSADAE
GAQLFAQSYNSAEQVLFQSVASNAHNTITAENARROEALISOEAAEAGKAK
ELYEPIMQNTDPPQRTIIGAVRTIGSANLPLAKROQYNALLSNRSRYSTKVCILN
KTAICWSLDPDLTNILASSRSYAMLLFMEGHNAAGLPKLFYEDFTALSNEAYKOD

KEYWORDS angiotensin-converting enzyme; peptidyl dipeptidase I.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS Soubrier, F.
TITLE NUCLEIC ACID CODING FOR THE HUMAN ANGIOGENSINE CONVERSION ENZYME
(ACE), AND ITS APPLICATIONS PARTICULARLY FOR THE IN VITRO DIAGNOSIS
OF HIGH BLOOD PRESSURE
JOURNAL Patent: WO 9003435-A 18 05-APR-1990;
Institut National de la Sante et de la Recherche Med
FEATURES
source
1. .4024
/organism="Homo sapiens"
/db_xref="taxon:9606"
23. .3943
/EC_number="3.4.15.1"
/codon_start=1
/product="peptidyl dipeptidase I"
/protein_id="CA00097.1"
/db_xref="GI:14653"
/translation="MGASGRGPGLLPPLLLPPLPPLALDPLGLQGNFSADAE
GAQLFAOSYNSAEOVLFOSYAASWADNTNTAENAROEBAALLSEFAWQOKAK
ELVEPIWQNFDPOLRRIIGAVRTLGASNLPLARQOYNALLSMSTRISTAKYCLPN
KTATCWSLDPLDNLIIASSRSYAMLLEFANEGMNAAGIPLKPLYEDFTALSNAYKOD
GFHTGAYWRSWYNSPTFEDDLHLVQOLEPLYNLHAFVRAALHRRYGDRIYINLRGP
IPALIGDMAOSWENTYDMVPPRDKPNLDVSTMLQGNATHMRVAFEEFTSLE
LSPMPPEWEGSMLEKPADGREVYCHASANDFNKRFRIKQCTRYMDOLSTYHNEM
GHIOYLYQYKDLVSLRGANPGFHEAIGDLAVSTPEHLKIGLDRTNDESD
INYLKMALEKIAFLPGGYLVDMQRMVFGSGRLVSPNEEDMWYLRKYOGICPPVR
NETHFDAGAKFHVNVNPIYIRYFVSFLQGFHALCKEAGEBPLQOICRSTKNG
AKLRKVLQAGSSRPMOEVLMKDWGLDIAQOLPLKYOPTOMLOBOONONGEVLGMP
EYOWHPRPLPNYREGIDLVNDEASAFVEYEDTSOVVWNEVLEAMNNTNTTET
SKILLQNMOTANHTLKYGTQARKFDVNOONTTKRIKKVODLERAPALPAQLEET
NKILLDEITYSVATVCHPNQSCQLDEPLINWATSRKEDLLAMEGMDKAGRAI
LOFPKXVELLINQARLVYDAGDSWRSMTPEPSLQDLERLEQELQPIYLNHAYV
RRALHRYGAQHINLEGPRIPAHLIGNMAQOTWSNITDLYVFPSPASMDTTEAMLKOG
WTPRRMKREADDEFTSGILLPVPEFPMKSMLEKTPGREGVYCHASANDFNKGRFRI
KQCTRYMDOLSTYHNEMGHIOYPMOYKQDLPVALREGANPGFHEAIGDLAVSTYH
HLSTNLSSBGSDEHDINFLKMAUDKIAFLPEPSLVDMQRMVFDGSLTKTNGY
HWSMLRKYOGICPPVRPTQDEDPGAKFHPSSVPYIRYFVSFLQGFHALCKEAGEBPLQOICRSTKNG
GHGTPKXVELLINQARLVYDAGDSWRSMTPEPSLQDLERLEQELQPIYLNHAYV
LDMLRTENELHGEKLGMPQYNMTPNSAISEGRLDGSRVSLGLDLDAQARVQGMILL
LFGLGIALVTFILGISORLFSIRHRSILRHSHGPFQSGSEVELRHS"

BASE COUNT 857 a 1263 c 1175 g 729 t

ORIGIN

Query Match 1.1%; Score 22; DB (c); Length 4024;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 922 CAGAGGCCAAGGAGCTGTATG 943
Db 317 CAGAGGCCAAGGAGCTGTATG 338
|||||
|||||

RESULT 29
AF118569 24070 bp DNA linear PRI 08-AUG-2000
LOCUS Homo sapiens angiotensin I converting enzyme precursor (DCP1) gene,
DEFINITION alternative splice products, complete cds.
ACCESSION AF118569
VERSION AF118569
KEYWORDS AF118569.1 GI:4732025
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE Rieder, M.J., Taylor, S.L., Clark, A.G. and Nickerson, D.A.
AUTHORS 1 (bases 1 to 24070)
TITLE Sequence variation in the human angiotensin converting enzyme
JOURNAL Nat. Genet. 22 (1), 59-62 (1999)

MEDLINE 99251580
PUBMED 10319862
REFERENCE 2 (bases 1 to 24070)
AUTHORS Rieder, M.J.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1999) Molecular Biotechnology, University of
Washington, UW Health Sciences Building, Rm. K316, 1705 NE Pacific,
Seattle, WA 98195, USA
FEATURES
source
1. .24070
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q23 between D17S794 and D17S795"
117
/frequency="0.05"
/replace="L"
334
/frequency="0.09"
/replace="L"
complement(<337. .>488)
/note="putative"
/rpt_family="L2"
/rpt_type-dispersed
/note="putative"
/rpt_family="MIR"
/rpt_type-dispersed
2400
/frequency="0.32"
/replace="L"
2547
/frequency="0.23"
/replace="C"
/gene="DCP1"
/note="Somatic form of ACE; dipeptidyl carboxypeptidase 1;
kininase II"
/codon_start=1
/product="angiotensin I converting enzyme precursor"
/protein_id="AAD28560.1"
/db_xref="GI:4732026"
/translation="MGASGRGPGLLPPLLLPPLPPLALDPLGLQGNFSADAE
GAQLFAOSYNSAEOVLFOSYAASWADNTNTAENAROEBAALLSEFAWQOKAK
ELVEPIWQNFDPOLRRIIGAVRTLGASNLPLARQOYNALLSMSTRISTAKYCLPN
KTATCWSLDPLDNLIIASSRSYAMLLEFANEGMNAAGIPLKPLYEDFTALSNAYKOD
GFHTGAYWRSWYNSPTFEDDLHLVQOLEPLYNLHAFVRAALHRRYGDRIYINLRGP
IPALIGDMAOSWENTYDMVPPRDKPNLDVSTMLQGNATHMRVAFEEFTSLE
LSPMPPEWEGSMLEKPADGREVYCHASANDFNKRFRIKQCTRYMDOLSTYHNEM
GHIOYLYQYKDLVSLRGANPGFHEAIGDLAVSTPEHLKIGLDRTNDESD
INYLKMALEKIAFLPGGYLVDMQRMVFGSGRLVSPNEEDMWYLRKYOGICPPVR
NETHFDAGAKFHVNVNPIYIRYFVSFLQGFHALCKEAGEBPLQOICRSTKNG
AKLRKVLQAGSSRPMOEVLMKDWGLDIAQOLPLKYOPTOMLOBOONONGEVLGMP
EYOWHPRPLPNYREGIDLVNDEASAFVEYEDTSOVVWNEVLEAMNNTNTTET
SKILLQNMOTANHTLKYGTQARKFDVNOONTTKRIKKVODLERAPALPAQLEET
NKILLDEITYSVATVCHPNQSCQLDEPLINWATSRKEDLLAMEGMDKAGRAI
LOFPKXVELLINQARLVYDAGDSWRSMTPEPSLQDLERLEQELQPIYLNHAYV
RRALHRYGAQHINLEGPRIPAHLIGNMAQOTWSNITDLYVFPSPASMDTTEAMLKOG


```

AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (23-APR-1999) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE    4 (bases 1 to 80858)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (03-JUN-1999) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE    5 (bases 1 to 80858)
AUTHORS      Waterston,R.
TITLE        Direct Submission
JOURNAL      Submitted (30-SEP-2000) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Sep 30, 2000 this sequence version replaced gi:5001526.

COMMENT      ----- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: saplens@watson.wustl.edu
              ----- Summary Statistics
              -----
              Center project name: H_NH0104K07
              -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanesi,J.J. and de Jong,P.-J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)

VECTOR: pBac3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-356B17; the clone sequenced
to the right is CTB-94H21, 200 bp overlap. Actual start of this
clone is at base position 1 of RP11-104K7; actual end is at base
position 91620 of CTB-94H21.

FEATURES
SOURCE
1. 80858
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2-q22"
/clone="RP11-104K7"
/clone_id="RPCI-11"
misc_feature
1..350

```

```
misc_feature
1. .336 /note="similar to EST A1360347 (NID:94111968) qy89g04.x1"
repeat_region
12. .176 /note="similar to EST T72671 (NID:9689346) yd19g06.s1"
repeat_region
350. .401 /rpt_family="Retroviral"
repeat_region
404. .461 /rpt_family="L1"
repeat_region
469. .488 /rpt_family="L1"
repeat_region
645. .820 /rpt_family="(TAA)n"
repeat_region
943. .1485 /rpt_family="L1"
repeat_region
1534. .1765 /rpt_family="Alu"
repeat_region
1768. .1856 /rpt_family="L1"
repeat_region
1875. .2301 /rpt_family="L1"
repeat_region
2684. .2936 /rpt_family="MER4-group"
repeat_region
2937. .3058 /rpt_family="MER4-group"
repeat_region
3059. .3357 /rpt_family="MER4-group"
repeat_region
3358. .3463 /rpt_family="Alu"
repeat_region
3790. .4000 /rpt_family="Alu"
repeat_region
4404. .4718 /rpt_family="L2"
repeat_region
4753. .4969 /rpt_family="Alu"
repeat_region
5669. .5716 /rpt_family="MIR"
repeat_region
7163. .7235 /rpt_family="MIR"
repeat_region
7293. .7331 /rpt_family="MIR"
repeat_region
7686. .7801 /rpt_family="(CA)n"
repeat_region
8278. .8321 /rpt_family="Alu"
repeat_region
8322. .8608 /rpt_family="MER21-group"
repeat_region
8609. .9272 /rpt_family="Alu"
repeat_region
9402. .9693 /rpt_family="MER21-group"
repeat_region
9697. .10140 /rpt_family="MER21-group"
repeat_region
10161. .10469 /rpt_family="L1"
repeat_region
10497. .10690 /rpt_family="Alu"
repeat_region
10810. .11053 /rpt_family="MER21-group"
repeat_region
11056. .11368 /rpt_family="MER4-group"
repeat_region
11714. .12015 /rpt_family="Alu"
repeat_region
12384. .12552 /rpt_family="Alu"
repeat_region
12553. .12849 /rpt_family="Alu"
repeat_region
12850. .12978 /rpt_family="Alu"
repeat_region
13354. .13553 /rpt_family="Alu"
repeat_region
13603. .13709 /rpt_family="MER4-group"
repeat_region
/rpt_family="L1"
```

```

repeat_region 13710..14005 /rpl_family="Alu"
repeat_region 14006..14247 /rpl_family="L1"
repeat_region 14377..14536 /rpl_family="MERL_type"
repeat_region 14599..14943 /rpl_family="MERL_type"
repeat_region 15193..15413 /rpl_family="MERL"
repeat_region 15569..15616 /rpl_family="L2"
repeat_region 15972..16262 /rpl_family="MERL_type"
repeat_region 16469..16591 /rpl_family="Alu"
repeat_region 17610..17739 /rpl_family="Alu"
repeat_region 18087..18387 /rpl_family="Alu"
repeat_region 18641..18659 /rpl_family="L1"
repeat_region 18660..18958 /rpl_family="L1"
repeat_region 18959..19194 /rpl_family="Alu"
repeat_region 19220..19495 /rpl_family="L1"
repeat_region 19597..19725 /rpl_family="Alu"
repeat_region 19726..20052 /rpl_family="L1"
repeat_region 20055..20225 /rpl_family="MERL"

```

```

Query Match 1.1% Score 22; DB 9; Length 80858;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1366 AAAAAAGATTGCACCTTATG 1387
Db 71233 AAAAAAGATTGCACCTTATG 71212

```

```

RESULT 31
AC103946 201500 bp DNA linear HTG 21-AUG-2002
DEFINITION Mus musculus clone RP23-85E10, WORKING DRAFT SEQUENCE, 9 ordered
ACCESSION AC103946
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULTOP.
SOURCE Mus musculus house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 201500)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 201500)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colanellano,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gilde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margis,N., Matthews,C.,
McCarthy,M., McKean,P., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
COMMENT

```

```

COMMENT

```

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Rella,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schpack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thoman,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 201500)
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barina,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepe,Y., Collymore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Linblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Rella,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schpack,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thoman,N., Stojanovic,N., Talamas,J.,
 Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced g1:19881886.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L19635
 Center clone name: 85_E.10
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 199978 bases at least Q40
 Consensus quality: 200513 bases at least Q30
 Consensus quality: 200620 bases at least Q20
 Insert size: 198000; agarose-fp
 Insert size: 200700; sum-of-ctnigs
 Quality coverage: 10.3 in Q20 bases; agarose-fp
 Quality coverage: 10.2 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 20114: contig of 20114 bp in length
 * 20115 20214: gap of 100 bp
 * 20215 24101: contig of 3887 bp in length
 * 24102 24201: gap of 100 bp
 * 24202 27267: contig of 3066 bp in length
 * 27268 27367: gap of 100 bp

```
* 27368 29341: contig of 1974 bp in length
* 29342 29441: gap of 100 bp
* 29442 33107: contig of 3666 bp in length
* 33108 33207: gap of 100 bp
* 33208 38389: contig of 5182 bp in length
* 38390 38489: gap of 100 bp
* 38490 129944: contig of 91455 bp in length
* 129945 130044: gap of 100 bp
* 130045 164792: contig of 34748 bp in length
* 164793 164892: gap of 100 bp
* 164893 201500: contig of 36608 bp in length.
FEATURES
    source
        1..201500
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone_id="RP23-85E10"
            /clone_lib="RPC1-23 Female Mouse BAC"
            1..20114
                /note="assembly-fragment"
                clone_end:SP6
                vector_side:left"
            20215..24101
                /note="assembly-fragment"
            24202..27267
                /note="assembly-fragment"
            27368..29341
                /note="assembly-fragment"
            29442..33107
                /note="assembly-fragment"
            33208..38389
                /note="assembly-fragment"
            38490..129944
                /note="assembly-fragment"
            130045..164792
                /note="assembly-fragment"
            164893..201500
                /note="assembly-fragment"
            clone_end:T7
            vector_side:right"
BASE COUNT  60676 a 39450 c 40269 g 60305 t 800 others
ORIGIN
Query Match      1.1%; Score 22; DB 2; Length 201500;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2020 TCTTCTCCTTTGAAGTTGTCA 2041
    |||||
Db 43062 TCTTCTCCTTTGAAGTTGTCA 43083

RESULT 32
LOCUS      E50434
DEFINITION Novel protein.
ACCESSION  E50434
VERSION    E50434.1 GI:18633519
KEYWORDS   JP 2001057890-A/2.
SOURCE     synthetic construct.
ORGANISM   artificial sequences.
REFERENCE  1 (bases 1 to 21)
AUTHORS    Omori,A. and Ito,M.
TITLE      Novel protein
JOURNAL    Patent: JP 2001057890-A 2 06-MAR-2001;
MITSUBISHI CHEM CORP
OS Artificial Sequence
PN JP 2001057890-A/2
PF 06-MAR-2001
PR 23-AUG-1999 JP 1999235218
PI AKIRA OMORI, MAKOTO ITO
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/80, PC
```

```
C12N15/00, C12N5/00
CC
FH key Location/Qualifiers
FT source 1..21
FT /Organism='Artificial Sequence'.
FEATURES
    source
        1..21
            /organism="synthetic construct"
            /db_xref="taxon:32630"
BASE COUNT  6 a 1 c 8 g 6 t
ORIGIN
Query Match      1.0%; Score 21; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 562 CTGGTGATTTGAATGAGAA 582
    |||||
Db 1 CTGGTGATTTGAATGAGAA 21

RESULT 33
LOCUS      AF201331
DEFINITION Rattus norvegicus strain F344/N angiotensin-converting enzyme (Ace)
ACCESSION  AF201331
VERSION    AF201331.1 GI:11493660
KEYWORDS   mRNA, complete cds.
SOURCE     Rattus norvegicus.
ORGANISM   Rattus norvegicus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 4014)
AUTHORS    Jafarian-Tehrani,M., Listwak,S., Barrientos,R.M., Michaud,A.,
            Corvol,P. and Sternberg,E.M.
TITLE      Characterization of a missense mutation in the angiotensin
            I-converting enzyme cDNA in exudative inflammation resistant F344/N
            rats
JOURNAL    Unpublished
AUTHORS    2 (bases 1 to 4014)
            Jafarian-Tehrani,M., Listwak,S., Barrientos,R.M., Michaud,A.,
            Corvol,P. and Sternberg,E.M.
TITLE      Direct Submission
JOURNAL    Submitted (02-NOV-1999) CNE, NIMH, 10 Center Dr., Bethesda, MD
            20892-1284, USA
FEATURES
    source
        1..4014
            Location/Qualifiers
            1..4014
                /organism="Rattus norvegicus"
                /strain="F344/N"
                /db_xref="taxon:10116"
                /chromosome="10"
                /tissue_type="lung"
            1..4014
                /gene="Ace"
            1..23
                /gene="Ace"
            24..3965
                /gene="Ace"
            /codon_start=1
            /product="angiotensin-converting enzyme"
            /protein_id="AA035596.1"
            /db_xref="GI:11493661"
            /translation="MGASAGGGRWPLSPPLMLSLLLLLPPSPAPALDGLQPN
            FSADGAGQLPADYSNNSAEVYMFQSTASVAHDTNITEENARLQEEALINQFAEV
            WGRKAKELYESIMQWFTDQKLRIIGSVOTGPAHLPLTORLQYNSLNNRIYSTG
            KVCPEPKATGCSIDPELNTLASRRAKYKLFAMEGWDHVGIPLRLYODPTALSN
            EAYRQDGSDDTGATYRSYTESPESIEHYHOVEPYLMLHAFYRALHRRGDKY
            INLGP1PAHLGDMQWQSMENITDMVVPEDKPLDVTSTVQKGNATHTMFAEES
            FETSLGSPMPPEFMAESMLEKPADREVVCASAMDEYNNKDRIRKQCTSTVMDQVS
            TVHEMGHVQYLYQKDLHVSILRGANGFHEALGDVLAISVSTPAHLAKTGLDRA
            NDIESDINYLKMLEKIAFLPFGYLVDQRMGVFSGRTPPSRYNIDMWYLRTRYGI
```


During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Mp: MOPREP; Information on the MOPREP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/MOPREP This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/MGP/Chr6>

RP11-134L4 is from the library RPCI-11.1 constructed by the group of Pletier de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-134L4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-134L4 is at 53423 in this sequence. The true right end of clone RP11-486M3 is at 100 in this sequence.

FEATURES

Source

1. 53423

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/clone="RP11-134L4"

/clone_id="RPCI-11.1"

/complement(1..114)

/note="match: GSS: Em:AQ132196"

/complement(1..94)

/note="match: GSS: Em:AQ277160"

304..612

/note="AluSg repeat: matches 1..308 of consensus"

756..947

/note="MER63A repeat: matches 5..202 of consensus"

1148..1243

/note="48 copies 2 mer tt 61% conserved"

1262..1458

/note="match: GSS: Em:AQ081950"

1344..1485

/note="FLAM_C repeat: matches 2..140 of consensus"

1490..1809

/note="AluSx repeat: matches 1..312 of consensus"

2477..3001

/note="L2 repeat: matches 2278..2735 of consensus"

3287..3477

/note="match: GSS: Em:B95534"

3482..3650

/note="match: GSS: Em:B95534"

3786..4147

/note="match: GSS: Em:AQ672311"

4156..4460

/note="AluSP repeat: matches 1..307 of consensus"

4588..4778

/note="MIR repeat: matches 2..208 of consensus"

5837..5991

/note="MIR repeat: matches 34..185 of consensus"

6610..6736

/note="L2 repeat: matches 2619..2746 of consensus"

7387..7785

/note="LINEC repeat: matches 272..671 of consensus"

repeat_region

7812..9031

/note="LINEC repeat: matches 1331..2435 of consensus"

repeat_region

9032..9342

/note="AluSx repeat: matches 1..309 of consensus"

repeat_region

9343..9656

/note="LINEC repeat: matches 2435..2741 of consensus"

misc_feature

complement(10314..10795)

repeat_region

10822..13811

/note="match: GSS: Em:AQ835745"

repeat_region

14336..14421

/note="LINEC repeat: matches 272..2803 of consensus"

repeat_region

14336..14421

/note="MIR repeat: matches 63..149 of consensus"

repeat_region

14347..14835

/note="AluY repeat: matches 1..289 of consensus"

misc_feature

complement(14845..15241)

misc_feature

15243..15683

/note="match: GSS: Em:AQ221492"

repeat_region

16434..16742

/note="AluSg repeat: matches 1..308 of consensus"

repeat_region

18333..18395

/note="L2 repeat: matches 2668..2731 of consensus"

repeat_region

18609..18842

/note="Aluub repeat: matches 1..234 of consensus"

repeat_region

18927..19219

/note="Aluub repeat: matches 1..292 of consensus"

repeat_region

19534..19673

/note="L2 repeat: matches 2224..2356 of consensus"

misc_feature

19890..20426

/note="match: GSS: Em:AQ081102"

misc_feature

19890..20412

/note="match: GSS: Em:AQ081070"

misc_feature

19890..20330

/note="match: GSS: Em:B88888"

misc_feature

19932..20397

/note="match: GSS: Em:AQ014577"

repeat_region

20610..21212

/note="match: GSS: Em:AQ787006"

repeat_region

20888..21119

/note="AluSx repeat: matches 1..228 of consensus"

repeat_region

21150..21435

/note="AluSx repeat: matches 1..300 of consensus"

repeat_region

21436..21600

/note="AluSg/x repeat: matches 134..297 of consensus"

repeat_region

21823..22102

/note="AluJo repeat: matches 1..282 of consensus"

repeat_region

22115..22306

/note="FLAM10 repeat: matches 5981..6322 of consensus"

repeat_region

24101..24483

/note="WSTB repeat: matches 1..426 of consensus"

misc_feature

complement(24849..25332)

repeat_region

25497..25665

/note="match: GSS: Em:AQ77587"

repeat_region

26517..26616

/note="L2 repeat: matches 2319..2488 of consensus"

repeat_region

26517..26616

/note="L2 repeat: matches 2500..2605 of consensus"

repeat_region

26617..26920

/note="AluSx repeat: matches 1..304 of consensus"

repeat_region

26921..27079

/note="L2 repeat: matches 2605..2735 of consensus"

repeat_region

27095..27181

/note="L2 repeat: matches 2656..2745 of consensus"

repeat_region

27774..27865

/note="46 copies 2 mer ca 69% conserved"

repeat_region

27941..28237

/note="AluSx repeat: matches 1..298 of consensus"

repeat_region

28385..28679

/note="Aluub repeat: matches 1..284 of consensus"

repeat_region

28875..29019

/note="MIR repeat: matches 48..195 of consensus"

repeat_region

30363..30458

/note="HERV140 repeat: matches 4..107 of consensus"

repeat_region

31215..31457

```

repeat_region /note="HERV140 repeat: matches 1368. .1617 of consensus"
31558. .31601
/note="HERV140 repeat: matches 2693. .2736 of consensus"
repeat_region 31760. .32072
/note="AluX repeat: matches 6. .310 of consensus"
32109. .32453
/note="LTR40a repeat: matches 106. .519 of consensus"
32537. .32582
/note="23 copies 2 mer tt 87% conserved"
repeat_region 35347. .35422
/note="MIR repeat: matches 59. .141 of consensus"
35386. .36273
/note="L1MC3 repeat: matches 7004. .7734 of consensus"
36240. .36605
/note="match: GSS: Em:AQ122318"
36318. .36585
/note="AluX repeat: matches 1. .267 of consensus"
36844. .36969
/note="L2 repeat: matches 2627. .2749 of consensus"
37410. .37670
/note="L2 repeat: matches 2470. .2747 of consensus"
37479. .38006
/note="match: GSS: Em:AQ16771"
38016. .38203
/note="L2 repeat: matches 2563. .2747 of consensus"
38417. .38452
/note="18 copies 2 mer aa 88% conserved"
38640. .38968
/note="Charliel repeat: matches 1. .2739 of consensus"
41104. .41158
/note="MIR repeat: matches 79. .135 of consensus"
41186. .41596
/note="match: GSS: Em:AQ199716"
41191. .41581
/note="match: GSS: Em:AQ546772"
41333. .41465
/note="MIR repeat: matches 5. .143 of consensus"
repeat_region
misc_feature
misc_feature
repeat_region

```

```

Query Match 1.0% Score 21: DB 9: Length 53423:
Best Local Similarity 100.0% Pred. No. 4.9:
Matches 21: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

```

Oy 1358 AGCGAATTAAAAAGAAATTG 1378
|||||
Db 22405 AGCGAATTAAAAAGAAATTG 22385

```

```

RESULT 38
AC131469 76700 bp DNA linear HTG 23-AUG-2002
LOCUS Rattus norvegicus clone CH230-193C19, *** SEQUENCING IN PROGRESS
DEFINITION *** 44 unordered pieces.
ACCESSION AC131469
VERSION AC131469.1 GI:22450473
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE 1 (bases 1 to 76700)
AUTHORS Murny,D,Marie, Metzker,M, Lee, A, Abranzone,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alsbrooks,S, Amth,A, Anguiano,D,
Aoyalebech,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Biswal,N, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryant,N, Buhay,C, Burch,P, Butrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Cessar,H, Center,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Devila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,
Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgievski,E., Geier,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulik,S., Hume,J., Idlebit,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowals,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Louisege,D., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindran,M., Mahmood,M., Mallory,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A.,
Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Mundasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwakoelamoh,O., Okunou,G.,
Olarpunsgoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poldexter,A.,
Popovic,D., Primus,E., Pu,L., L., Puzo,M., Quiroz,J., Rachlin,E.,
Reeves,K., Regier,M.A., Reigh,R., Kelly,B., Kelly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatsman,S., Shen,H., Shetty,J., Shwartsbeyn,A.,
Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
Song,X., Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willison,R., Wleczyk,R.,
Wooden,H., Morley,K., Wright,J., Wright,R., Wu,J., Yakub,S.,
Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhausen,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.
Unpublished
2 (bases 1 to 76700)
Rat Genome Sequencing Consortium.
Submitted (23-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: CGNT
Center clone name: CH230-193C19
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap version 0.990329
Consensus quality: 29743 bases at least Q40
Consensus quality: 32480 bases at least Q20
Consensus quality: 34360 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1006: contig of 1006 bp in length
* 1007 1106: gap of unknown length
* 1107 2204: contig of 1098 bp in length

```

2205 2304: gap of unknown length
2305 3320: contig of 1016 bp in length
2321 3420: gap of unknown length
3421 4474: contig of 1054 bp in length
4475 4574: gap of unknown length
4575 5704: contig of 1130 bp in length
5705 5804: gap of unknown length
5805 7210: contig of 1406 bp in length
7211 7310: gap of unknown length
7311 8442: contig of 1132 bp in length
8443 8542: gap of unknown length
8543 10195: contig of 1653 bp in length
10196 10285: gap of unknown length
10286 11430: contig of 1195 bp in length
11431 11590: gap of unknown length
11591 13046: contig of 1456 bp in length
13047 13146: gap of unknown length
13147 14505: contig of 1359 bp in length
14506 14605: gap of unknown length
14606 16138: contig of 1533 bp in length
16139 16238: gap of unknown length
16239 17782: contig of 1544 bp in length
17783 17882: gap of unknown length
17883 19585: contig of 1713 bp in length
19586 19685: gap of unknown length
19686 20790: contig of 1095 bp in length
20791 20890: gap of unknown length
20891 22384: contig of 1494 bp in length
22385 22484: gap of unknown length
22485 23511: contig of 1027 bp in length
23512 23611: gap of unknown length
23612 25171: contig of 1560 bp in length
25172 25271: gap of unknown length
25272 27046: contig of 1775 bp in length
27047 27146: gap of unknown length
27147 28748: contig of 1602 bp in length
28749 28848: gap of unknown length
28849 29995: contig of 1147 bp in length
29996 30095: gap of unknown length
30096 31696: contig of 1601 bp in length
31697 31997: gap of unknown length
31998 32947: contig of 1151 bp in length
32948 33047: gap of unknown length
33049 34144: contig of 1097 bp in length
34145 34244: gap of unknown length
34245 35729: contig of 1485 bp in length
35730 35829: gap of unknown length
35830 37855: contig of 2026 bp in length
37856 37955: gap of unknown length
37956 39379: contig of 1424 bp in length
39380 39479: gap of unknown length
39480 41067: contig of 1588 bp in length
41068 41167: gap of unknown length
41169 42304: contig of 1137 bp in length
42305 42404: gap of unknown length
42406 44115: contig of 1711 bp in length
44116 44215: gap of unknown length
44217 45401: contig of 1186 bp in length
45402 45501: gap of unknown length
45502 47109: contig of 1608 bp in length
47110 47209: gap of unknown length
47211 50085: contig of 2876 bp in length
50086 50185: gap of unknown length
50186 53328: contig of 3143 bp in length
53329 53428: gap of unknown length
53429 55854: contig of 2426 bp in length
55855 55954: gap of unknown length
55955 58755: contig of 2801 bp in length
58756 58855: gap of unknown length
58856 60750: contig of 1895 bp in length
60751 60850: gap of unknown length
60851 62682: contig of 1832 bp in length
62683 62782: gap of unknown length

62783 64194: contig of 1412 bp in length
64195 64294: gap of unknown length
64295 64949: contig of 1655 bp in length
64949 65950: gap of unknown length
65950 66049: gap of unknown length
66049 68333: contig of 2284 bp in length
68333 68433: gap of unknown length
68433 70841: contig of 2408 bp in length
70841 70941: gap of unknown length
70941 73306: contig of 2365 bp in length
73306 73406: gap of unknown length
73406 76700: contig of 3294 bp in length.
Location/Qualifiers
* 73407
FEATURES
source 1..76700
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-193C19"
BASE COUNT 20693 a 15464 c 14004 g 21005 t 5534 others
ORIGIN
Query Match 1.0%; Score 21; DB 2; Length 76700;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2021 CTTCTCCTTTTGAAGTTGCA 2041
Db 14318 CTTCTCCTTTTGAAGTTGCA 14338
RESULT 39
AC096435
LOCUS
DEFINITION
AC096435
Rattus norvegicus clone CH230-44G5. *** SEQUENCING IN PROGRESS ***
60 unordered pieces.
AC096435
AC096435.3 GI:21723831
VERSION
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
ORGANISM
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 100858)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbarta,J., Benton,J., Bimaga,K., Blankensburg,K., Bonnin,D.,
Boudet,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doutwalte,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Huijk,S., Hunne,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolyet,S., Joudah,S.,
Karlssohn,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabac,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,N., Newton,J., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokoko,S., Ogih,M., Okunou,G.,
Oraguine,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Saverly,G.,

Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, S.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 100858)
Worley, K.C.
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 100858)
Worley, K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:17946751.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: CH230-44G5
Center clone name: GF0E
----- Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.930329
Consensus quality: 62832 bases at least Q40
Consensus quality: 67157 bases at least Q30
Consensus quality: 69957 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1547: contig of 1547 bp in length
1548: gap of unknown length
1548: contig of 1016 bp in length
2664: gap of unknown length
2763: gap of unknown length
3984: contig of 1221 bp in length
3985: gap of unknown length
4084: gap of unknown length
5098: contig of 1014 bp in length
5198: gap of unknown length
5199: gap of unknown length
6258: contig of 1060 bp in length
6259: gap of unknown length
6358: gap of unknown length
7430: contig of 1072 bp in length
7431: gap of unknown length
8692: contig of 1162 bp in length
8693: gap of unknown length
8792: gap of unknown length
8793: contig of 1434 bp in length
10227: gap of unknown length
10326: gap of unknown length
10327: contig of 1024 bp in length
11350: gap of unknown length
11351: gap of unknown length
11450: gap of unknown length
12456: contig of 1006 bp in length
12457: gap of unknown length
12556: gap of unknown length
13585: contig of 1029 bp in length
13586: gap of unknown length
13586: contig of 1102 bp in length
14787: gap of unknown length
14787: contig of 1254 bp in length
14888: gap of unknown length
16141: gap of unknown length
16241: gap of unknown length
16242: contig of 1384 bp in length
17626: gap of unknown length
17725: gap of unknown length
17726: gap of unknown length
18929: contig of 1204 bp in length
18930: gap of unknown length
19029: gap of unknown length
19300: contig of 1473 bp in length
20502: gap of unknown length
20503: gap of unknown length
20603: contig of 1282 bp in length
21855: gap of unknown length
21955: gap of unknown length
23092: contig of 1137 bp in length
23191: gap of unknown length
23192: gap of unknown length
25216: contig of 2025 bp in length
25217: gap of unknown length
25317: gap of unknown length
26837: contig of 1520 bp in length
26936: gap of unknown length
26937: gap of unknown length
28478: contig of 1542 bp in length
28479: gap of unknown length
28579: gap of unknown length
29821: contig of 1242 bp in length
29920: gap of unknown length
29921: gap of unknown length
31496: contig of 1576 bp in length
31596: gap of unknown length
31597: gap of unknown length
32778: contig of 1182 bp in length
32779: gap of unknown length
32878: gap of unknown length
33909: contig of 1031 bp in length
33910: gap of unknown length
34009: gap of unknown length
34010: gap of unknown length
35555: contig of 1546 bp in length
35556: gap of unknown length
35556: gap of unknown length
36804: contig of 1149 bp in length
36805: gap of unknown length
36904: gap of unknown length
38329: contig of 1425 bp in length
38429: gap of unknown length
38430: gap of unknown length
38430: contig of 1181 bp in length
39611: gap of unknown length
39710: gap of unknown length
39711: gap of unknown length
40731: contig of 1021 bp in length
40732: gap of unknown length
40831: gap of unknown length
42285: contig of 1454 bp in length
42286: gap of unknown length
42385: gap of unknown length
42386: contig of 1512 bp in length
43898: gap of unknown length
43898: gap of unknown length
43997: gap of unknown length
45096: contig of 1099 bp in length
45097: gap of unknown length
45196: gap of unknown length
47290: contig of 2094 bp in length
47291: gap of unknown length
47390: gap of unknown length
47391: gap of unknown length
48592: contig of 1202 bp in length
48593: gap of unknown length
48692: gap of unknown length
50756: contig of 2064 bp in length
48693: gap of unknown length
50757: gap of unknown length
50856: gap of unknown length
52145: contig of 1289 bp in length
52146: gap of unknown length
52245: gap of unknown length
53392: contig of 1147 bp in length
53393: gap of unknown length
53492: gap of unknown length
55013: contig of 1521 bp in length
55014: gap of unknown length
55113: gap of unknown length
55114: gap of unknown length
56969: contig of 1856 bp in length
56970: gap of unknown length
57069: gap of unknown length
58886: contig of 1817 bp in length
58887: gap of unknown length
58987: gap of unknown length
60202: contig of 1216 bp in length
60203: gap of unknown length
60302: gap of unknown length
61902: contig of 1600 bp in length
61903: gap of unknown length
62003: gap of unknown length
63404: contig of 1402 bp in length
63405: gap of unknown length
63504: gap of unknown length
63505: gap of unknown length
64786: contig of 1282 bp in length
64787: gap of unknown length
64886: gap of unknown length
64887: gap of unknown length
64887: contig of 1568 bp in length
66455: gap of unknown length
66554: gap of unknown length
67793: contig of 1239 bp in length
67794: gap of unknown length
67894: gap of unknown length
70504: contig of 2611 bp in length
70505: gap of unknown length

Query Match	1.0%:	Score 21:	DB 2:	Length 100858:
Best Local Similarity	100.0%:	Pred. No. 4.9:		
Matches	21:	Conservative	0:	Mismatches
			0:	Indels
			0:	Gaps
OY	923	AGAAGGCCAAGAGCTGTATG	943	
Db	3188	AGAAGGCCCAAGAGCTGTATG	3208	

RESULT 40	197743 bp	DNA	linear	HTG 31-JUL-2002
AC098290/c				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
AC098290	197743 bp	DNA	linear	HTG 31-JUL-2002
Rattus norvegicus clone CH230-1K15,				
72 unordered pieces.				
AC098290				
AC098290.3	GI:21953793			
HTG: HTGS_PHASE1.				
Norway rat.				
Rattus norvegicus				
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
Rattus.				
1 (bases 1 to 197743)				
Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-Osman,F.R., Allen,C.,				
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayela,M., Banks,T.,				
Barbarta,J., Benton,J., Bimage,K., Blankenburg,K., Bonnh,D.,				
Bouck,J., Burch,P., Bivela,M., Brown,M., Brown,M., Bryan,N.P.,				
Bunay,C., Burket,C., Burrell,K.L., Byrd,N.C.,				
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,				
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,				
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,				
David,M.L., Davis,C., Davy-Carroll,L., Deederich,D.A.,				
Delaney,K.R., Delgado,O., Dem,A.L., Ding,X., Dinh,H.H.,				
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,				
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,				
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frintz,P.,				
Gabisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,				
Gorell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,				
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,				
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,				
Homsil,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,				
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,				
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,				
Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,				
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,M., Louisgeed,H.,				
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,				
Maheeshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,				
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,				
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,				
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,				
Nguyen,N., Nickerson,E., Nwokwoko,S., Oguh,M., Okunolu,G.,				
Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,				
Peters,L., Pickens,R., Plimus,E., Pu,L.L., Qulies,M., Ren,Y.,				
Rives,M., Rojas,A., Rojoubokan,I., Rolfe,M., Ruiz,S., Savery,G.,				
Scheerer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,				
Soderstrom,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,				
Sutton,A., Swalek,A., Tabori,P., Tamefisa,A., Tamefisa,K., Tang,H.,				
Taney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,				
Uemari,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,				
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlingon,S.,				
Williams,G., Williamson,A., Wlarczyk,R., Wooden,S., Worley,K.,				

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 19743)
AUTHORS	Worley, R.C.
TITLE	Direct Submission
JOURNAL	Submitted (23-OCT-2001)
REFERENCE	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
AUTHORS	3 (bases 1 to 179743)
TITLE	Worley, R.C.
JOURNAL	Direct Submission
REFERENCE	Submitted (31-JUL-2002)
AUTHORS	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Jul 24, 2002 this sequence version replaced gi:20976531.

On Jul 24, 2002 this sequence version replaced g1:20976531.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: TUGA
Center clone name: CH230-1K15
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 131319 bases at least Q40
Consensus quality: 137136 bases at least Q30
Consensus quality: 141601 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
NOTE: This is a 'working draft' sequence. It currently consists of 72 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	1138:	contig of 1138 bp in length
1139	1238:	gap of unknown length
1239	2301:	contig of 1063 bp in length
2302	2401:	gap of unknown length
2402	3496:	contig of 1095 bp in length
3497	3596:	gap of unknown length
3597	5069:	contig of 1473 bp in length
5070	5169:	gap of unknown length
5170	6811:	contig of 1642 bp in length
6812	6911:	gap of unknown length
6912	8070:	contig of 1159 bp in length
8071	8170:	gap of unknown length
8171	9299:	contig of 1129 bp in length
9300	9399:	gap of unknown length
9400	10818:	contig of 1419 bp in length
10819	10918:	gap of unknown length
10919	12095:	contig of 1177 bp in length
12096	12195:	gap of unknown length
12196	13277:	contig of 1082 bp in length
13278	13377:	gap of unknown length
13378	14438:	contig of 1061 bp in length
14439	14538:	gap of unknown length
14539	16006:	contig of 1468 bp in length
16007	16106:	gap of unknown length
16107	18128:	contig of 2022 bp in length
18129	18628:	gap of unknown length
18629	19776:	contig of 1448 bp in length
19777	21436:	gap of unknown length
21437	21536:	contig of 1660 bp in length
21536:	gap of unknown length	

```

* 21537 22548: contig of 1012 bp in length
* 22549 22648: gap of unknown length
* 22649 24318: contig of 1670 bp in length
* 24319 24418: gap of unknown length
* 24419 25580: contig of 1162 bp in length
* 25581 25680: gap of unknown length
* 25681 27093: contig of 1413 bp in length
* 27094 27193: gap of unknown length
* 27194 28470: contig of 1277 bp in length
* 28471 28570: gap of unknown length
* 28571 30684: contig of 2114 bp in length
* 30685 30784: gap of unknown length
* 30785 31851: contig of 1067 bp in length
* 31852 31951: gap of unknown length
* 31952 33328: contig of 1377 bp in length
* 33329 33428: gap of unknown length
* 33429 35336: contig of 1908 bp in length
* 35337 35436: gap of unknown length
* 35437 37015: contig of 1579 bp in length
* 37016 37115: gap of unknown length
* 37116 38430: contig of 1315 bp in length
* 38431 38530: gap of unknown length
* 38531 40403: contig of 1873 bp in length
* 40404 40503: gap of unknown length
* 40504 41571: contig of 1068 bp in length
* 41572 41671: gap of unknown length
* 41672 43207: contig of 1536 bp in length
* 43208 43307: gap of unknown length
* 43308 43175: contig of 1868 bp in length
* 43176 45275: gap of unknown length
* 45276 47055: contig of 1780 bp in length
* 47056 47155: gap of unknown length
* 47156 49525: contig of 2370 bp in length
* 49526 49625: gap of unknown length
* 49626 51279: contig of 1654 bp in length
* 51280 51379: gap of unknown length
* 51380 53441: contig of 2062 bp in length
* 53442 53541: gap of unknown length
* 53542 55028: contig of 1487 bp in length
* 55029 55128: gap of unknown length
* 55129 56806: contig of 1678 bp in length
* 56807 56906: gap of unknown length
* 56907 58657: contig of 1751 bp in length
* 58658 60008: contig of 1251 bp in length
* 60009 60108: gap of unknown length
* 60109 62324: contig of 2216 bp in length
* 62325 62424: gap of unknown length
* 62425 64265: contig of 1841 bp in length
* 64266 64365: gap of unknown length
* 64366 65759: contig of 1394 bp in length
* 65760 65859: gap of unknown length
* 65860 67579: contig of 1720 bp in length
* 67580 67679: gap of unknown length
* 67680 69047: contig of 1368 bp in length
* 69048 69147: gap of unknown length
* 69148 71865: contig of 2718 bp in length
* 71866 71965: gap of unknown length
* 71966 75130: contig of 3165 bp in length
* 75131 75230: gap of unknown length
* 75231 78455: contig of 3225 bp in length
* 78456 78555: gap of unknown length
* 78556 80136: contig of 1581 bp in length
* 80137 80236: gap of unknown length
* 80237 82812: contig of 2576 bp in length
* 82813 82912: gap of unknown length
* 82913 84864: contig of 1952 bp in length
* 84865 87080: gap of unknown length
* 84965 87180: contig of 2116 bp in length
* 87081 87181: gap of unknown length
* 87181 89315: contig of 2035 bp in length
* 89316 92002: contig of 2687 bp in length

```

```

* 92003 92102: gap of unknown length
* 92103 94959: contig of 2857 bp in length
* 94960 95059: gap of unknown length
* 95060 97782: contig of 2723 bp in length

Query Match 1.0%; Score 21; DB 2; Length 179743;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1220 TGCTTCACAGTGGAGAGCTGA 1240
Db 44323 TGCTTCACAGTGGAGAGCTGA 44303

RESULT 41
AC068982
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
REFERENCE
TITLE
JOURNAL
COMMENT

AC068982 181663 bp DNA linear HTG 07-JUL-2000
Homo sapiens chromosome 6 clone RP11-79506, WORKING DRAFT SEQUENCE,
7 unordered pieces.
AC068982
AC068982.3 GI:8099316
HTG, HTGS_PHASE1, HTGS_DRAFT.
Homo sapiens.
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 181663)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 181663)
Waterston, R.H.
Direct Submission
Submitted (15-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 27, 2000 this sequence version replaced gi:7924013.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0795006
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: Plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175821 bases at least Q40
Consensus quality: 177817 bases at least Q30
Consensus quality: 178903 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 181063; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 6.15 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2688: contig of 2688 bp in length
* 2689 2788: gap of unknown length
* 2789 9166: contig of 6378 bp in length
* 9167 9266: gap of unknown length
* 9267 17646: contig of 8380 bp in length
* 17647 17746: gap of unknown length
* 17747 32787: contig of 15041 bp in length

```

FEATURES	source	location/Qualifiers
*	32788	32887: gap of unknown length
*	32888	43393: contig of 16506 bp in length
*	49394	49493: gap of unknown length
*	96494	96672: contig of 50179 bp in length
*	99673	99772: gap of unknown length
*	99773	181663: contig of 81891 bp in length.
FEATURES		
source		
misc_feature	1. 181663	/organism="Homo sapiens"
misc_feature	/db_xref="taxon:9606"	
misc_feature	/chromosome="6"	
misc_feature	/clone="RP11-79506"	
misc_feature	1. 2688	
misc_feature	/note="assembly_name:Contig8"	
misc_feature	2789. 9166	
misc_feature	/note="assembly_name:Contig9	
misc_feature	clone_end:77	
misc_feature	vector_slide:left"	
misc_feature	9267. 17646	
misc_feature	/note="assembly_name:Contig10"	
misc_feature	17747. 32787	
misc_feature	/note="assembly_name:Contig11"	
misc_feature	32888. 49393	
misc_feature	/note="assembly_name:Contig12"	
misc_feature	49494. 99672	
misc_feature	/note="assembly_name:Contig13"	
misc_feature	99773. 181663	
misc_feature	/note="assembly_name:Contig14"	
BASE COUNT	52019 a 37585 c 38060 g 53390 t	609 others
ORIGIN		
Query Match	1.0% Score 21: DB 2: Length 181663:	
Best Local Similarity	100.0%: Prid. No. 5:	
Matches	21: Conservative 0: Mismatches	0: Indels 0: Gaps 0:
OY	1358 AGCATTAAAGAAATTTG 1378	
Db	116879 AGCCATTAAAAAGAAATTTG 116899	
RESULT 42		
AC120772/c		
LOCUS		
DEFINITION	AC120772 186651 bp DNA linear HTG 24-JUL-2002	
ACCESSION	Rattus norvegicus clone CH230-203F10.*** SEQUENCING IN PROGRESS	
VERSION	***, 63 unordered pieces.	
KEYWORDS	AC120772	
SOURCE	AC120772.3 GI:21909159	
ORGANISM	HTG: HTGS. PHASE1.	
	Rattus norvegicus.	
	Rattus norvegicus.	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
	Rattus.	
REFERENCE	1 (bases 1 to 186651)	
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,	
	Alsbrooks,S.U., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,	
	Barbata,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,	
	Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,	
	Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,	
	Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,	
	Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,	
	Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,	
	Dayla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,	
	DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,	
	Douthett,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,	
	Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,	
	Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,	
	Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,	
	Gorell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,	
	Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J.,	
	Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,	
	Homsli,F., Howard,S., Huber,B., Huliy,S., Hume,J., Jackson,L.E.,	
	Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,	

REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL
Kratzson, E., Kelly, S., Khan, U., King, L., Koryavch, J., Kovar, C., Kravtsov, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisedge, R., Lozada, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Matindale, A., Martinez, E., Massey, E., Maunley, E., McLeod, M. P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S., Ogun, M., Okwonn, G., Ogunyeye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Plinius, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherzer, S., Scott, G., Shen, H., Shoochari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svalek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umanai, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S., Williams, G., Williamson, A., Wleciyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.	Unpublished 2 (bases 1 to 166651) Worley, K.C. Direct Submission Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 166651) Worley, K.C. Direct Submission Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 19, 2002 this sequence version replaced gi:20564394.	Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu Project Information Center project name: GYGI Center clone name: CH230-203P10 Sequencing vector: Plasmid Chemistry: Dye-terminator Big Dye: 100% of reads Assembly program: Phrap: version 0.990329 Consensus quality: 142024 bases at least Q40 Consensus quality: 148489 bases at least Q30 Consensus quality: 152361 bases at least Q20	NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html) NOTE: This is a "working draft" sequence. It currently consists of 63 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. 1 1113: contig of 1113 bp in length 1114: gap of unknown length 1213: gap of unknown length 1214: contig of 1275 bp in length 2488: gap of unknown length 2589: gap of unknown length 2589: contig of 1469 bp in length 4057: gap of unknown length 4157: gap of unknown length 5351: contig of 1194 bp in length 5352: gap of unknown length 5451: gap of unknown length 7014: contig of 1563 bp in length 7015: gap of unknown length 7114: gap of unknown length 8812: contig of 1698 bp in length

```

* 8813 8912: gap of unknown length
* 8913 9958: contig of 1046 bp in length
* 9959 10058: gap of unknown length
* 10059 11113: contig of 1055 bp in length
* 11114 11213: gap of unknown length
* 11214 12318: contig of 1105 bp in length
* 12319 12418: gap of unknown length
* 12419 14029: contig of 1611 bp in length
* 14030 14129: gap of unknown length
* 14130 15613: contig of 1484 bp in length
* 15614 15713: gap of unknown length
* 15714 16851: contig of 1138 bp in length
* 16852 16951: gap of unknown length
* 16952 18585: contig of 1634 bp in length
* 18586 18685: gap of unknown length
* 18686 20981: contig of 2296 bp in length
* 20982 21081: gap of unknown length
* 21082 23190: contig of 2109 bp in length
* 23191 23290: gap of unknown length
* 23291 24804: contig of 1514 bp in length
* 24805 24904: gap of unknown length
* 24905 26496: contig of 1592 bp in length
* 26497 26596: gap of unknown length
* 26597 28532: contig of 1936 bp in length
* 28533 28632: gap of unknown length
* 28633 30042: contig of 1410 bp in length
* 30043 30142: gap of unknown length
* 30143 31313: contig of 1171 bp in length
* 31314 31413: gap of unknown length
* 31414 32881: contig of 1468 bp in length
* 32882 32981: gap of unknown length
* 32982 34460: contig of 1479 bp in length
* 34461 34560: gap of unknown length
* 34561 36685: contig of 2125 bp in length
* 36686 36785: gap of unknown length
* 36786 38279: contig of 1494 bp in length
* 38280 38379: gap of unknown length
* 38380 39958: contig of 1579 bp in length
* 39959 40058: gap of unknown length
* 40059 41667: contig of 1709 bp in length
* 41668 41867: gap of unknown length
* 41868 44854: contig of 2987 bp in length
* 44855 44954: gap of unknown length
* 44955 46927: contig of 1973 bp in length
* 46928 47027: gap of unknown length
* 47028 48811: contig of 1784 bp in length
* 48812 48911: gap of unknown length
* 48912 50907: contig of 1996 bp in length
* 50908 51007: gap of unknown length
* 51008 54318: contig of 3311 bp in length
* 54319 54418: gap of unknown length
* 54419 55922: contig of 1504 bp in length
* 55923 56022: gap of unknown length
* 56023 58319: contig of 2297 bp in length
* 58320 58419: gap of unknown length
* 58420 60899: contig of 2480 bp in length
* 60900 60999: gap of unknown length
* 61000 63820: contig of 2821 bp in length
* 63821 63920: gap of unknown length
* 63921 66860: contig of 2940 bp in length
* 66861 66960: gap of unknown length
* 66961 68743: contig of 1783 bp in length
* 68744 68843: gap of unknown length
* 68844 71603: contig of 2760 bp in length
* 71604 71703: gap of unknown length
* 71704 74491: contig of 2788 bp in length
* 74492 74591: gap of unknown length
* 74592 77206: contig of 2615 bp in length
* 77207 77306: gap of unknown length
* 77307 80580: contig of 3274 bp in length
* 80581 80680: gap of unknown length
* 80681 83678: contig of 2998 bp in length
* 83679 83778: gap of unknown length

```

```

* 83779 86973: contig of 3195 bp in length
* 86974 87073: gap of unknown length
* 87074 90192: contig of 3119 bp in length
* 90193 90292: gap of unknown length
* 90293 95917: gap of unknown length
* 95917 96016: gap of unknown length
* 96017 100345: contig of 4329 bp in length
* 100346 100445: gap of unknown length
* 100446 104770: contig of 4325 bp in length
* 104771 104870: gap of unknown length
* 104871 110000: contig of 5130 bp in length
* 110001 110100: gap of unknown length
* 110101 113116: contig of 3016 bp in length
* 113117 113216: gap of unknown length
* 113217 117352: contig of 4136 bp in length
* 117353 117452: gap of unknown length
* 117453 121355: contig of 3903 bp in length
* 121356 121455: gap of unknown length
* 121456 124627: contig of 3172 bp in length
* 124628 124727: gap of unknown length
* 124728 130091: contig of 5364 bp in length
* 130092 130191: gap of unknown length
* 130192 135015: contig of 4824 bp in length

Query Match
Best Local Similarity 100.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1268 TTGTGATGTCAGATTGTTA 1288
Db 143561 TTGTGATGTCAGATTGTTA 143541

RESULT 43
AC106294
LOCUS
DEFINITION
Rattus norvegicus clone CH230-76010, *** SEQUENCING IN PROGRESS
AC106294
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Rattus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 191192)
Muzny,D.M., Adams,C., Adio-Odnola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimagne,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Boyle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Bunyah,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Howski,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoen,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,J., Liu,W., Louseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Martinez,E.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Matinez,E.,
Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,

```

Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunola, G.,
Oragunye, N., Oyedero, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Plums, E., Pu, L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojupokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,
Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Umani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 191192)

Worley, K.C.

Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 191192)

Worley, K.C.

Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:18138815.

----- Genome Center -----

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center Project name: GK0M
Center Clone name: CH230-76010
----- Summary Statistics -----
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 125322 bases at least Q40
Consensus quality: 132642 bases at least Q30
Consensus quality: 138789 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1388: contig of 1388 bp in length
* 1389 1488: gap of unknown length
* 1489 2669: contig of 1181 bp in length
* 2670 2769: gap of unknown length
* 2770 3808: contig of 1039 bp in length
* 3809 3908: gap of unknown length
* 3909 4929: contig of 1021 bp in length
* 4930 5029: gap of unknown length
* 5030 6223: contig of 1194 bp in length
* 6224 6323: gap of unknown length
* 6324 7626: contig of 1303 bp in length
* 7627 7726: gap of unknown length
* 7727 8871: contig of 1145 bp in length
* 8872 8971: gap of unknown length
* 8972 10196: contig of 1225 bp in length
* 10197 10296: gap of unknown length
* 10297 11381: contig of 1085 bp in length
* 11382 11481: gap of unknown length

11482 13166: contig of 1685 bp in length
* 13167 13266: gap of unknown length
* 13267 14374: contig of 1108 bp in length
* 14375 14474: gap of unknown length
* 14475 15707: contig of 1233 bp in length
* 15708 15807: gap of unknown length
* 15808 16862: contig of 1055 bp in length
* 16863 16962: gap of unknown length
* 16963 18060: contig of 1098 bp in length
* 18061 18160: gap of unknown length
* 18161 19511: contig of 1351 bp in length
* 19512 19612: gap of unknown length
* 19613 20795: contig of 1183 bp in length
* 20796 20894: gap of unknown length
* 20895 22725: contig of 1831 bp in length
* 22726 22825: gap of unknown length
* 22826 23832: contig of 1007 bp in length
* 23833 23932: gap of unknown length
* 23933 25112: contig of 1180 bp in length
* 25113 25212: gap of unknown length
* 25213 26435: contig of 1223 bp in length
* 26436 26535: gap of unknown length
* 26536 28046: contig of 1511 bp in length
* 28047 28146: gap of unknown length
* 28147 30041: contig of 1895 bp in length
* 30042 30141: gap of unknown length
* 30142 32406: contig of 2265 bp in length
* 32407 32506: gap of unknown length
* 32507 33667: contig of 1161 bp in length
* 33668 33767: gap of unknown length
* 33768 35177: contig of 1410 bp in length
* 35178 35277: gap of unknown length
* 35278 37020: contig of 1743 bp in length
* 37021 37120: gap of unknown length
* 37121 39350: contig of 2230 bp in length
* 39351 39450: gap of unknown length
* 39451 41003: contig of 1553 bp in length
* 41004 41103: gap of unknown length
* 41104 42456: contig of 1353 bp in length
* 42457 42556: gap of unknown length
* 42557 44708: contig of 2152 bp in length
* 44709 44808: gap of unknown length
* 44809 45863: contig of 1055 bp in length
* 45864 45963: gap of unknown length
* 45964 47550: contig of 1587 bp in length
* 47551 47650: gap of unknown length
* 47651 49436: contig of 1786 bp in length
* 49437 49536: gap of unknown length
* 49537 51357: contig of 1821 bp in length
* 51358 51457: gap of unknown length
* 51458 54021: contig of 2564 bp in length
* 54022 54121: gap of unknown length
* 54122 55686: contig of 2465 bp in length
* 55687 56686: gap of unknown length
* 56687 58137: contig of 1451 bp in length
* 58138 58237: gap of unknown length
* 58237 60536: contig of 2299 bp in length
* 60537 60636: gap of unknown length
* 60637 63842: contig of 3206 bp in length
* 63843 63942: gap of unknown length
* 63943 65950: contig of 2008 bp in length
* 65951 66050: gap of unknown length
* 66051 67766: contig of 1716 bp in length
* 67767 67866: gap of unknown length
* 67867 69616: contig of 1750 bp in length
* 69617 69716: gap of unknown length
* 69717 71766: contig of 2050 bp in length
* 71767 71866: gap of unknown length
* 71867 74342: contig of 2476 bp in length
* 74343 74442: gap of unknown length
* 74443 75742: contig of 1300 bp in length
* 75743 75842: gap of unknown length
* 75843 78294: contig of 2452 bp in length

```

* 78295 78394: gap of unknown length
* 78395 78952: contig of 1558 bp in length
* 79953 80052: gap of unknown length
* 80053 83228: contig of 3176 bp in length
* 83229 83328: gap of unknown length
* 83329 87055: contig of 3727 bp in length
* 87056 87155: gap of unknown length
* 87156 90022: contig of 2867 bp in length
* 90023 90122: gap of unknown length
* 90123 92263: contig of 2141 bp in length
* 92264 92363: gap of unknown length
* 92364 95642: contig of 3279 bp in length
* 95643 95742: gap of unknown length
* 95743 97352: contig of 1610 bp in length
* 97353 97452: gap of unknown length
* 97453 99828: contig of 2376 bp in length

Query Match 1.0%; Score 21; DB 2; Length 191192;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 938 TGTATGCTCTGCTCCAGC 958
Db 113832 TGTATGCTCTGCTCCAGC 113852

RESULT 44
AC094442/C AC094442 197174 bp DNA linear HTG 10-JUL-2002
LOCUS Rattus norvegicus clone CH230-4F5, *** SEQUENCING IN PROGRESS ***
DEFINITION 70 unordered pieces.
ACCESSION AC094442.3 GI:21716483
VERSION AC094442
KEYWORDS HTG: HTGS, PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 197174)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alabrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Blinze,K., Blankenburg,K., Bonini,D.,
Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burreli,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,K.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,D., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Koirah,J., Kovar,C.,
Kritovoy,C., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosier,N., Nickerson,E., Nwokenwo,S., Ogun,M., Okunolu,G.,
Ongunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rojle,M., Ruiz,S., Savery,G.,
Scheerer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,

```

```

TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Teliford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 197174)
Worley,K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 197174)
Worley,K.C.
Direct Submission
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 9, 2002 this sequence version replaced gi:17941170.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAPP
Center clone name: CH230-4F5
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 158750 bases at least Q40
Consensus quality: 163771 bases at least Q30
Consensus quality: 168185 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1093: contig of 1093 bp in length
* 1094 1193: gap of unknown length
* 1194 2298: contig of 1105 bp in length
* 2299 2398: gap of unknown length
* 2399 4247: contig of 1849 bp in length
* 4248 4347: gap of unknown length
* 4348 5913: contig of 1566 bp in length
* 5914 6013: gap of unknown length
* 6014 7073: contig of 1060 bp in length
* 7074 7173: gap of unknown length
* 7174 8519: contig of 1346 bp in length
* 8520 8619: gap of unknown length
* 8620 9903: contig of 1284 bp in length
* 9904 10003: gap of unknown length
* 10004 11377: contig of 1374 bp in length
* 11378 11477: gap of unknown length
* 11478 12943: contig of 1466 bp in length
* 12944 13043: gap of unknown length
* 13044 14531: contig of 1488 bp in length
* 14532 14631: gap of unknown length
* 14632 16736: contig of 2105 bp in length
* 16737 16836: gap of unknown length
* 16837 18076: contig of 1240 bp in length
* 18077 18176: gap of unknown length
* 18177 19513: contig of 1337 bp in length

```

```

* 19514 19513: gap of unknown length
* 19614 20718: contig of 1105 bp in length
* 20719 20818: gap of unknown length
* 20819 22279: contig of 1461 bp in length
* 22280 22379: gap of unknown length
* 22380 24578: contig of 2199 bp in length
* 24579 24678: gap of unknown length
* 24679 26727: contig of 2049 bp in length
* 26728 26827: gap of unknown length
* 26828 28392: contig of 1465 bp in length
* 28293 29900: contig of 1508 bp in length
* 29901 30000: gap of unknown length
* 30001 31343: contig of 1343 bp in length
* 31344 33191: contig of 1748 bp in length
* 33192 33291: gap of unknown length
* 33292 35020: contig of 1729 bp in length
* 35021 35120: gap of unknown length
* 35121 36637: contig of 1517 bp in length
* 36638 36737: gap of unknown length
* 36738 38139: contig of 1402 bp in length
* 38140 38239: gap of unknown length
* 38240 40561: contig of 2322 bp in length
* 40562 40661: gap of unknown length
* 42493 42593: contig of 1832 bp in length
* 42594 44303: contig of 1710 bp in length
* 44304 44403: gap of unknown length
* 44404 46466: contig of 2063 bp in length
* 46467 46566: gap of unknown length
* 46567 48004: contig of 1438 bp in length
* 48005 48104: gap of unknown length
* 48105 50303: contig of 2199 bp in length
* 50304 53233: gap of unknown length
* 53234 53333: gap of unknown length
* 53334 55265: contig of 1932 bp in length
* 55266 55365: gap of unknown length
* 55366 57056: contig of 1691 bp in length
* 57057 57156: gap of unknown length
* 57157 59608: contig of 2452 bp in length
* 59609 59708: gap of unknown length
* 59709 63931: contig of 4223 bp in length
* 63932 64031: gap of unknown length
* 64032 66878: contig of 2847 bp in length
* 66879 66978: gap of unknown length
* 66979 69578: contig of 2600 bp in length
* 69579 71920: gap of unknown length
* 71921 72020: gap of unknown length
* 72021 74778: contig of 2758 bp in length
* 74779 74878: gap of unknown length
* 74879 77904: contig of 3026 bp in length
* 77905 78004: gap of unknown length
* 78005 81637: contig of 3633 bp in length
* 81638 81737: gap of unknown length
* 81738 83825: contig of 2088 bp in length
* 83826 83925: gap of unknown length
* 83926 86560: contig of 2635 bp in length
* 86561 86660: gap of unknown length
* 86661 88789: contig of 2129 bp in length
* 88790 88889: gap of unknown length
* 88890 90661: contig of 1772 bp in length
* 90662 90761: gap of unknown length
* 90762 94063: contig of 3302 bp in length
* 94064 94163: gap of unknown length
* 94164 97840: contig of 3677 bp in length
* 97841 97940: gap of unknown length
* 97941 100540: contig of 2600 bp in length
* 100541 100640: gap of unknown length
* 100641 103145: contig of 2505 bp in length
* 103146 103245: gap of unknown length

```

```

* 103246 105812: contig of 2567 bp in length
* 105813 105912: gap of unknown length
* 105913 109869: contig of 3957 bp in length
* 109870 109970: gap of unknown length
* 109970 113238: contig of 3269 bp in length
* 113239 113338: gap of unknown length
* 113339 116891: contig of 3553 bp in length
* 116892 116991: gap of unknown length
* 116992 120117: contig of 3126 bp in length

Query Match      1.0%: Score 21: DB 2: Length 197174:
Best Local Similarity 100.0%: Pred. No. 5:
Matches 21: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Qy      1220 TGCTTCACAGTGAGAGCTCA 1240
Db      161075 TGCTTCACAGTGAGAGCTCA 161055

RESULT 45
AC096705
LOCUS
DEFINITION
Rattus norvegicus clone CH230-218, *** SEQUENCING IN PROGRESS ***,
70 unordered pieces.
AC096705
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 212255)
REFERENCE
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barberia,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkelt,C., Burrell,K.L., Byrd,N.C.,
Carion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doutwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., Kling,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lounseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Merlindade,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunou,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,T., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Tusmanl,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wlecezyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

```



```

JOURNAL
TITLE
Direct Submission
Unpublished
2 (bases 1 to 212255)
REFERENCE
AUTHORS
Worley,K.C.
JOURNAL
TITLE
Direct Submission
Submitted (23-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE
AUTHORS
Worley,K.C.
JOURNAL
TITLE
Direct Submission
Submitted (31-Jul-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 24, 2002 this sequence version replaced gi:15799569.
COMMENT
-----
Center: Baylor College of Medicine
Genome Center
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: TUSJ
Center clone name: CH230-218
-----
Summary Statistics
Sequencing vector: Plasmid,
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 154890 bases at least Q40
Consensus quality: 161468 bases at least Q30
Consensus quality: 166127 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
* 1122: contig of 1122 bp in length
* 1123 1222: gap of unknown length
* 1223 2420: contig of 1198 bp in length
* 2420 2520: gap of unknown length
* 2521 3521: contig of 1001 bp in length
* 3522 3621: gap of unknown length
* 3622 4871: contig of 1250 bp in length
* 4872 4971: gap of unknown length
* 4972 6044: contig of 1073 bp in length
* 6045 6145: gap of unknown length
* 6145 7592: contig of 1448 bp in length
* 7593 7692: gap of unknown length
* 7693 8942: contig of 1250 bp in length
* 8943 9042: gap of unknown length
* 9043 10562: contig of 1520 bp in length
* 10563 10662: gap of unknown length
* 10663 11687: contig of 1024 bp in length
* 11687 11786: gap of unknown length
* 11787 13151: contig of 1365 bp in length
* 13152 13251: gap of unknown length
* 13252 14683: contig of 1432 bp in length
* 14684 14783: gap of unknown length
* 14784 15811: contig of 1028 bp in length
* 15812 15911: gap of unknown length
* 15912 17238: contig of 1327 bp in length
* 17239 17338: gap of unknown length
* 17339 18470: contig of 1132 bp in length
* 18471 18570: gap of unknown length
* 18571 19730: contig of 1160 bp in length
* 19731 19830: gap of unknown length
* 19831 20908: contig of 1078 bp in length
* 20909 21008: gap of unknown length

```

*	21009	22440:	contig of 1432	bp	in length
*	22441	22540:	gap of unknown	length	
*	22541	24410:	contig of 1870	bp	in length
*	24411	24510:	gap of unknown	length	
*	24511	25900:	contig of 1390	bp	in length
*	25901	26000:	gap of unknown	length	
*	26001	27471:	contig of 1471	bp	in length
*	27472	27571:	gap of unknown	length	
*	27572	29135:	contig of 1564	bp	in length
*	29136	29235:	gap of unknown	length	
*	29236	31590:	contig of 2355	bp	in length
*	31591	31690:	gap of unknown	length	
*	31691	32914:	contig of 1224	bp	in length
*	32915	33014:	gap of unknown	length	
*	33015	34913:	contig of 1899	bp	in length
*	34914	35013:	gap of unknown	length	
*	35014	36934:	contig of 1921	bp	in length
*	36935	37034:	gap of unknown	length	
*	37035	38412:	contig of 1378	bp	in length
*	38413	38512:	gap of unknown	length	
*	38513	39754:	contig of 1242	bp	in length
*	39755	39854:	gap of unknown	length	
*	39855	41115:	contig of 1261	bp	in length
*	41116	41215:	gap of unknown	length	
*	41216	42990:	contig of 1275	bp	in length
*	42491	42590:	gap of unknown	length	
*	42591	44367:	contig of 1777	bp	in length
*	44368	44467:	gap of unknown	length	
*	44468	45114:	contig of 1647	bp	in length
*	45115	46314:	gap of unknown	length	
*	46315	47602:	contig of 1288	bp	in length
*	47503	47602:	gap of unknown	length	
*	47603	50193:	contig of 2591	bp	in length
*	50194	50293:	gap of unknown	length	
*	50294	53097:	contig of 2804	bp	in length
*	53098	53197:	gap of unknown	length	
*	53198	55497:	contig of 2300	bp	in length
*	55498	55597:	gap of unknown	length	
*	55598	56971:	contig of 1374	bp	in length
*	56972	57071:	gap of unknown	length	
*	57072	59081:	contig of 2010	bp	in length
*	59082	59181:	gap of unknown	length	
*	59182	61500:	contig of 2469	bp	in length
*	61501	61750:	gap of unknown	length	
*	61751	63634:	contig of 1884	bp	in length
*	63635	63734:	gap of unknown	length	
*	63735	65382:	contig of 1548	bp	in length
*	65383	65482:	gap of unknown	length	
*	65483	67112:	contig of 1930	bp	in length
*	67113	67412:	gap of unknown	length	
*	67413	69137:	contig of 2325	bp	in length
*	69138	69637:	gap of unknown	length	
*	69638	72845:	contig of 3008	bp	in length
*	72846	72945:	gap of unknown	length	
*	72946	75984:	contig of 3039	bp	in length
*	75985	76084:	gap of unknown	length	
*	76085	78508:	contig of 2424	bp	in length
*	78509	78608:	gap of unknown	length	
*	78609	81574:	contig of 2966	bp	in length
*	81575	81574:	gap of unknown	length	
*	81675	84191:	contig of 2517	bp	in length
*	84192	84291:	gap of unknown	length	
*	84292	86433:	contig of 2142	bp	in length
*	86434	86534:	gap of unknown	length	
*	86534	86549:	contig of 3116	bp	in length
*	86550	89749:	gap of unknown	length	
*	89750	93795:	contig of 4046	bp	in length
*	93796	93895:	gap of unknown	length	
*	93896	97729:	contig of 3834	bp	in length
*	97730	97829:	gap of unknown	length	
*	97830	101606:	contig of 3777	bp	in length
*	101607	101706:	gap of unknown	length	
*	101707	104900:	contig of 3194	bp	in length

Query Match	Best Local Similarity	1.0%; Score 21; DB 2; Length 212255;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY 923 AGAAGCCAGGAGCTGATG 943		
DB 167067 AGAAGCCAGGAGCTGATG 167087		
RESULT 46		
LOCUS AL592564	AL592564	750 bp DNA
DEFINITION Meleagris gallopavo DNA; STS TC0252, sequence tagged site.	AL592564	linear STS 16-JUL-2001
ACCESSION AL592564		
VERSION AL592564.1 GI:14715682		
KEYWORDS STS.		
SOURCE turkey		
ORGANISM Meleagris gallopavo		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Archosauria; Aves; Neognathae; Galliformes; Meleagrididae;		
Meleagris.		
1 (bases 1 to 750)		
Morrice,D.		
Unpublished		
2 (bases 1 to 750)		
Morrice,D.		
Direct Submission		
Submitted (16-JUL-2001) Morrice D., Roslin Institute, Roslin,		
Midlothian, EH25 9PS, UK david.morrice@bsrc.ac.uk		
Vector pBLUESCRIPT SK+		
R. Site 1 EcorI		
R. Site 2 EcorI		
Contact: David Morrice		
Dept. Genomics and Bioinformatics		
Roslin Institute, Roslin		
Midlothian, EH25 9PS, UK		
Tel: +44 (0) 131 527 4200		
Fax: +44 (0) 131 440 0434		
Email: david.morrice@bsrc.ac.uk		
Seq primer 17.		
Location/Qualifiers		
1..750		
/organism="Meleagris gallopavo"		
/db_xref="taxon:9103"		
1..750		
/standard_name="TC0252"		
BASE COUNT 215 a 163 c 143 g 207 t 22 others		
ORIGIN		
Query Match	1.0%; Score 20; DB 11; Length 750;	
Best Local Similarity 100.0%; Pred. No. 18;		
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY 275 TGAGTGCATTACACACAC 294		
DB 424 TGAGTGCATTACACACAC 443		
RESULT 47		
LOCUS HSN800587		
DEFINITION Homo sapiens mRNA; cDNA DKFZp564J1762 (from clone DKFZp564J1762).		
ACCESSION AL080086		
VERSION AL080086.1 GI:5262502		
KEYWORDS		
SOURCE Homo sapiens.		
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
1 (bases 1 to 1065)		

AUTHORS	Blum, H., Baunersachs, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUN-1999) MIPS, Am Klopferstritz 18a, D-82152 Martinsried, Germany
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; Sequenced by LMU (Ludwig Maximilians University, Munich/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp564J1762) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/ .
FEATURES	Location/Qualifiers 1..1065 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="DKFZp564J1762" /tissue_type="brain" /clone_id="564 (synonym: hfbz2). Vector: pAMPI; host X1-2blue; sites NotI + SalI" /dev_stage="fetal"
BASE COUNT	polyA_site 275 a 268 c 311 g 211 t
ORIGIN	
Query Match	1.0%; Score 20; DB 9; Length 1065;
Best Local Similarity	100.0%; Pred. No. 18;
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1575 CACATGAGCAGTGTCCCG 1554 121 CACATGAGCAGTGTCCCG 140
Db	
RESULT 48	
AF231711/c	1680 bp mRNA linear VRT 01-MAR-2000
LOCUS	
DEFINITION	Xenopus laevis 7-transmembrane receptor frizzled-1 mRNA, complete cds.
ACCESSION	AF231711
VERSION	AF231711.1 GI:7110523
KEYWORDS	
SOURCE	Xenopus laevis. Xenopus laevis. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Eupipidae; Xenopodinae; Xenopus.
ORGANISM	
REFERENCE	1 (bases 1 to 1680) Brown, J.D., Hallagan, S.E., McGrew, L.L., Miller, J.R. and Moon, R.T.
AUTHORS	
TITLE	The maternal Xenopus beta-catenin signaling pathway, activated by frizzled homologs, induces goosecoid in a cell non-autonomous manner
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1680) Brown, J.D., Hallagan, S.E. and Moon, R.T.
AUTHORS	
TITLE	Direct Submission
JOURNAL	Submitted (04-FEB-2000) Pharmacology and HHMI, University of Washington, Seattle, WA 98195, USA
FEATURES	Location/Qualifiers 1..1680 /organism="Xenopus laevis" /db_xref="taxon:8355" 1..1680 /codon_start=1 /product="7-transmembrane receptor frizzled-1" /protein_id="AA36979.1" /db_xref="GI:7110524" /translation="MKSHLQRCSAQCTRGSSLLISLVCISVEGYNGEKGISIPDHGYCOPISLPCDIAVNOTIMPLDHTNEDAGLEHORYPLVKQSPSEKFLFCGIVAPCTVEQALPCRCSDCARROCEALNMKFGQWPEBSLCEKFPINGAECLGQNTTESGTPPAVEPWTNSKRIYVYDKRCPCPALVPAIVNTHFGKDCGAP

gene
complement(2964. .3692)
/gene="gldb_L"
/note="gldb: The second copy of gldb gene is located in
the right arm of LADR"
complement(2964. .3692)
/gene="gldb_L"
/codon_start=1
/transl_table=4
/product="glucose inhibited division protein B"
/protein_id="AAF36756.1"
/db_xref="GI:7109685"
/translation="MLKIKDSLCKFNLTLDKOEDIAFFLEEIHSNOLFNTGYK
TELLAEMGVYILLASQSLVYSFNSNLIDGTGIGLIKIKIIPOLANYLVD
NAKKITPNEVIKYKLKFTNGVPALISREDNPFLKYGCPDVSQAWSIALNEL
GTOLKINGOIIHRSRDYQEIEFAKKNLSDLGLAFNNLIHYOPNSIFLVNYTKRKA
ALAPOKPREMSKIRLEIDDKAH"
complement(3685. .5520)
/gene="glda_L"
/note="glda: The second copy of glda gene is located in
the right arm of LADR"
complement(3685. .5520)
/gene="glda_L"
/codon_start=1
/transl_table=4
/product="glucose inhibited division protein A"
/protein_id="AAF36757.1"
/db_xref="GI:7109686"
/translation="MNRKFVVAGAGHGLEAAFIISKLNKRYLCVLDRKYVANCPRN
SVGGAKCIVIREIDALGIOALADSYALORILINSKGPGVCLRFQIDRVYKKMK
FLEDONDENNEELVEGEVTAEVKNDPTATGWIMBGKRLGAVIITFGYLKLFPS
KDVAENDEPEGRKSNNSLMSEPKVGPFILIRLKGTGPRIKSDSIDVSNLOIEGNCTE
EIFYHRSKNKRTIDELPCYLIHTHEERHKIINDNLHSANSGNITGVCRTPSIE
DKIVEFPKHQPIPLEPESLELTVYLGGFSTDISDVOKIIRLLGPLKRAEVIRK
GAIEYDAINPIQLPSLESKLKVNLPFPAQINGSGYEAGAOGGLAGIANOKIKR
KEPLISDDEAYEIVGMIDDIYTKGTDYRLTSSREALLANDNVLDPLIQCYEII
GISRKSDILYNQNLEKNKRLIEFLKDKVAMTYTLRAHTNNTNESYLEFLKBEIKI
IELKILEDVSNDELKLNKEITVEXKEGIIKRESRPVNSLKSISIKIPODIYDX
VONSIEDILINKIKKPILNALAQRISSINADIISLKTILEQNA"
5771..6259
/note="ORF162.L: The second copy of ORF162 CDS is located
in the right arm of LADR"
/codon_start=1
/transl_table=4
/product="unknown"
/protein_id="AAF36758.1"
/db_xref="GI:7109687"
/translation="MKLSLKSGLTREOKLERISKLTDELRITWIIILFGYLFVVMNA
FLIIDVACNFETYPNAESTLSWAGSOSFFPKNGSITTAATNMWTIFFRCIGSFVAG
WIGHGHKKKVLYTMGLMALSPFPIVAVYPPGGNNALVLSNQIIFYQGQDLSGGS
LT"

CDS
6369. .7349
/note="ORF326.L: The second copy of ORF326 CDS is located
in the right arm of LADR"
/codon_start=1
/transl_table=4
/product="unknown"
/protein_id="AAF36759.1"
/db_xref="GI:7109688"
/translation="MKRTTLLSVINSFCFNIGAASFSLGTFYSMTIKMEAVKSWYVIC
AAFLIIFATFLIYPIEGKETVLPKERTKMEPEPTIMSTLKNKNAVIMKCMJGSAWMLY
AVWALTCTSNSYIGTSPIYVNGKLFPHVCGTSKIAPYLGLVAGLFIAPPGKRTRYBR
SRFLKTOYSLGFEVYVAVLGFPGVANNICILAIGEIIIASFLAGIPLMGJLQVLLPI
HEFKASSKSCQFOGIIWGCVMLYTSDSLIAYSIESPFARKSVYDSINPASITA
FAVYIYISMENVLLIYLPKSGRIYNGEWBVLDTDKMPMSYNFYKGDIFK"
complement(7414. .8028)
/gene="def_L"
/note="def: The second copy of def gene is located in the
right arm of LADR"
complement(7414. .8028)
/gene="def_L"
/EC_number="3.5.1.31"
/codon_start=1
/transl_table=4

CDS

```

Query Match      1.0%; Score 20; DB 1; Length 17411;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1265 ATATTGTTGATGTGCAGATT 1284
        |||||||
Db       1556 ATATTGTTGATGTGCAGATT 1575

RESULT 50
AC116980
LOCUS
DEFINITION
AC116980      23461 bp      DNA      linear      HTG 05-APR-2002
Dictyostelium discoideum chromosome 2 map 1408084-1432089 strain
AA4, *** SEQUENCING IN PROGRESS ***, In ordered pieces.
AC116980
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Lehmann R., Baumgart C., Parra G., April J.F., Gulgo R., Kumpf K.,
Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A. and
Noegel A.A.
Sequence and Analysis of Chromosome 2 of Dictyostelium
Unpublished
The Dictyostelium Genome Sequencing Consortium
2 (bases 1 to 23461)
Baumgart C.
Direct Submission
Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COS predictions from Genaid may contain errors. Further information
is available from IMB Jena, Department of Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..23461
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44699"
/chromosome="2"
/map="1408084-1432089"
complement(2324..2590)
/note="ORF_ID:dd_01691"
/codon_start=1
/product="hypotheical protein"
/protein_id="AA08754.1"
/db_xref="GI:20042937"
/translation="MPCVILENPLSEPRIFCTNDLVEYNHESQEGHYILETF
KGRFKNSHLITITFENNCGGGSYKANYITLTNLRIVYKOTNKE"
join(2792..2868,2908..3049)
/note="ORF_ID:dd_01692"
/codon_start=1
/product="hypotheical protein"
/protein_id="AA08755.1"
/db_xref="GI:20042937"
/translation="MNAICIDAYFNMFKILTTQYTTIDELLLSLISLFEFVYVDD
DYUUDSGSNHNHNHNHNHNHNHNHI"
complement(join(3380..5292,5378..5641,5910..5937))
/note="ORF_ID:dd_01693"
/codon_start=1
/product="hypotheical protein"
FEATURES
source
CDS
CDS
CDS

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 09:08:54 ; Search time 338 Seconds
(without alignments)
9462.933 Million cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttcaagtgtcactactatg.....ttgaagtgtcactactag 2049

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by change to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	431.8	21.1	1194	10	US-09-739-907-49 Sequence 49, Appl
2	430.2	21.0	1222	10	US-09-739-907-31 Sequence 31, Appl
3	121.2	5.9	1941	10	US-09-777-710A-2 Sequence 2, Appl
4	121.2	5.9	2013	10	US-09-777-710A-16 Sequence 16, Appl
5	72.2	3.5	272	10	US-09-878-574-9145 Sequence 9145, Ap
6	39.8	1.9	713	10	US-09-910-943-91 Sequence 91, Appl
7	37.4	1.8	462	10	US-09-954-456-1057 Sequence 1057, Ap
8	36	1.7	170	10	US-09-878-574-8039 Sequence 8039, Ap
9	34	1.7	14155	9	US-10-108-605-102 Sequence 102, App
10	33.6	1.6	1898	10	US-09-822-849A-310 Sequence 310, App
11	33.4	1.6	415	9	US-10-186-846-2496 Sequence 2496, Ap
12	33.2	1.6	1137	9	US-10-184-644-518 Sequence 518, App
13	33.2	1.6	1137	9	US-10-184-634-518 Sequence 518, App
14	33	1.6	1523	9	US-10-184-644-290 Sequence 290, App
15	33	1.6	1523	9	US-10-184-634-290 Sequence 290, App
16	33	1.6	7040	9	US-10-172-086-13 Sequence 13, Appl
17	32.8	1.6	912	9	US-09-738-626-924 Sequence 924, Appl
18	32.8	1.6	1473	9	US-09-738-626-923 Sequence 923, App
19	32.8	1.6	2643	10	US-09-815-242-7116 Sequence 7116, Ap

20	32.8	1.6	4929	9	US-09-738-626-3394 Sequence 3394, Ap
21	32.8	1.6	8354	9	US-10-125-818-1 Sequence 1, Appl
22	32.8	1.6	1503841	9	US-09-946-807-1 Sequence 1, Appl
23	32.8	1.6	1503841	10	US-09-795-668-1 Sequence 1, Appl
24	32.8	1.6	1503841	10	US-09-795-668-1 Sequence 1, Appl
25	32.8	1.6	1830121	9	US-10-329-960-1 Sequence 1, Appl
26	32.8	1.6	3309400	9	US-09-738-626-1 Sequence 1, Appl
27	32.8	1.6	3309400	9	US-09-738-626-1 Sequence 1, Appl
28	32.6	1.6	471	9	US-10-066-543-2224 Sequence 2224, Ap
29	32.6	1.6	1199	12	US-10-044-090-9 Sequence 9, Appl
30	32.6	1.6	1242	9	US-10-083-357-601 Sequence 601, App
31	32.6	1.6	2393	10	US-09-964-824A-281 Sequence 281, App
32	32.6	1.6	3316	9	US-10-242-056-58 Sequence 58, Appl
33	32.6	1.6	3752	9	US-10-098-841-55 Sequence 55, Appl
34	32.6	1.6	3752	9	US-10-037-270-691 Sequence 691, App
35	32.6	1.6	7287	9	US-09-938-842A-1019 Sequence 1019, Ap
36	32.4	1.6	20300	9	US-09-939-209A-3 Sequence 3, Appl
37	32.4	1.6	20556	10	US-09-880-107-3945 Sequence 3945, Ap
38	32.2	1.6	547	9	US-09-918-995-27398 Sequence 27398, A
39	32.2	1.6	1570	9	US-09-938-842A-4436 Sequence 4436, Ap
40	32.2	1.6	19553	9	US-10-092-154-1425 Sequence 1425, Ap
41	32.2	1.6	19553	10	US-09-764-847-1425 Sequence 1425, Ap
42	32.2	1.6	70768	9	US-10-135-322-13 Sequence 13, Appl
43	32	1.6	238	10	US-09-878-574-10358 Sequence 10358, A
44	32	1.6	277	10	US-09-878-574-14108 Sequence 14108, A
45	32	1.6	383	10	US-09-878-574-1049 Sequence 1049, Ap

ALIGNMENTS

RESULT 1

US-09-739-907-49
Sequence 49, Application US/09739907
Patent No. US20010012889A1

GENERAL INFORMATION:
APPLICANT: Rosen et al.

TITLE OF INVENTION: 36 Human Secreted Proteins
FILE REFERENCE: P2022p1

CURRENT APPLICATION NUMBER: US/09/739, 907
CURRENT FILING DATE: 2000-12-20

PRIOR APPLICATION NUMBER: 09/348, 457
PRIOR FILING DATE: 1999-07-07

PRIOR APPLICATION NUMBER: 60/070, 567
PRIOR FILING DATE: 1998-01-07

PRIOR APPLICATION NUMBER: 60/070, 692
PRIOR FILING DATE: 1998-01-07

PRIOR APPLICATION NUMBER: 60/070, 704
PRIOR FILING DATE: 1998-01-07

PRIOR APPLICATION NUMBER: 60/070, 658
PRIOR FILING DATE: 1998-01-07

NUMBER OF SEQ ID NOS: 196
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 49

LENGTH: 1194

TYPE: DNA

ORGANISM: Homo sapiens

US-09-739-907-49

Query Match

Best Local Similarity 81.1%;

Matches 502; Conservative

Score 431.8; DB 10;

Length 1194;

Pred. No. 2,5e-132;

Mismatches 117; Indels 0;

Gaps 0;

1	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 3394, Ap
2	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
3	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
4	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
5	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
6	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
7	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
8	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
9	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
10	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
11	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
12	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
13	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
14	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
15	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
16	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
17	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
18	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
19	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
20	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
21	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
22	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
23	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
24	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
25	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
26	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
27	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
28	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
29	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
30	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
31	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
32	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
33	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
34	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
35	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
36	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
37	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
38	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
39	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
40	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
41	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
42	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
43	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
44	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl
45	TTCACTGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC	60	Sequence 1, Appl

Db 546 AGTCGTGCTTCAATCATGCGAAGAACTGATGGGTCATGCAACAGTGTTCAGCATC 605
Qy 181 GAACATGATGATTTCCCAACGACTGAGGTGGAGGTCCCTGAAGAGACTAGAGATAA 240
Db 606 GACATGGCATGGATGCCAAGGCTCAGGCTGGAGGTCTTGAAACACACTGCAGAGATAA 665
Qy 241 TATGGCTCTGTATGACAGACAAATGTTATCTGTAGTCCATTCAACACACTGTGGC 300
Db 666 TATGGCTCCCTGTACAGAAAGATTAATGTCATCTGTAGTGGCACTCACACTTTCAGGT 725
Qy 301 CCAGCAGGGTTTTCCAAATATACACTATATCTGCGCAGCGAGGATTCAGACACGG 360
Db 726 CCGGACGAGATTTTCCAGTATACCGGTGTTGTAATTCAGTGAAGGATTTAGCAATCA 785
Qy 361 ACCTTCAGTACATAGTCTGTGGATCATGAGAGCATGATATAGCTCACACAAATCTT 420
Db 786 ACTTTCAGCATATGTCACACTGGATCTTGAAGAGCATTTGACATACACAAATATG 845
Qy 421 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTCTAATGTGCAGATCAACCGAGC 480
Db 846 AAACGAGCAAAATCTTCATCAATAAAGAAATGTGATGTGTGCAGATCAACAGAGT 905
Qy 481 CCTCTCTTACCTTCTGAAATCCAGATGACAGAGACAGATTCCTCAACACAGAC 540
Db 906 CCGTATTTCTTACCTTCAAAATCCGCACTCAGAGAGACAGATTTCTTCAATACAGAC 965
Qy 541 AAGGAAATGCTGTGTTGAAAGTGTGATTTGATGAGAGACATTTGGGCTTATCAGC 600
Db 966 AAGGAAATGATAGTTTGAAGAAATGTGATTTGATGATGAGATGACTTGGGCTTATCAGT 1025
Qy 601 TGGTTGCCATCCACCCG 619
Db 1026 TTTTCATTCAGCAAGTCTG 1044

RESULT 2

US-09-739-907-31
; Sequence 31, Application US/09739907
; Patent No. US20010012889A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P2022P1
; CURRENT APPLICATION NUMBER: US/09/739,907
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/348,457
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1222
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-739-907-31

Query Match 21.0%; Score 430.2; DB 10; Length 1222;
Best Local Similarity 80.9%; Pred. No. 8.6e-132;
Matches 501; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 1 TTGAGTGCTACTACATTTGGGAGAGCGGATTCACAGACAGTGTCAATATC 60
Db 449 TTGAGTGCTACTACATTTGGGAGAGCGGATTCACAGACAGTGTCAATATC 508
Qy 61 AATTGATGGGCTATGGCAAAATGGCCAAATGACAGGGGTCTCCACACAGGCTTTC 120
|||||

Db 509 AATTGATGGGCTATGGCAAAATCCGGCAGAAATGACACAGGCGATCTCCACAGGCTATAC 568
Qy 121 AGCGTGCTTTTATCTTGGGAGATCCAGATGGGTCAATATGCAATGGCATTTGTGAGCGTG 180
Db 569 AGTGTGCTTTCATTCATGGGAGAAACCTGATGGGTCCATATGAAACAGTGTTCAGCATC 628
Qy 181 GAACATGATGATTTCCCAACGACTGAGGTGGAGGTCTTGAAAGACTAGAGATAA 240
Db 629 GACATPAGCATGGATTCACAAAGGCTCAGGCTGGAGGTCTTGAAACACACTGCAGATAA 688
Qy 241 TATGGCTCTGTATGACAGACAAATGTTATCTGTAGTGGCACTTCACACACTGTGGC 300
Db 689 TATGGCTCCCTGTACAGAAAGATTAATGTCATCTGTAGTGGCACTTCACACTTTCAGGT 748
Qy 301 CCAGCAGGGTTTTCCAAATATACACTATATCTGCGCAGCGAGGATTCAGACACCGG 360
Db 749 CCGGACGAGATTTTCCAGTATACCGGTGTTGTAATTCAGTGAAGGATTTAGCAATCA 808
Qy 361 ACCTTCAGTACATAGTCTGTGGATCATGAGAGCATGATATAGCTCACACAAATCTT 420
Db 809 ACTTTCAGCATATGTCACACTGGATCTTGAAGAGCATTTGACATACACAAATATG 868
Qy 421 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTCTAATGTGCAGATCAACCGAGC 480
Db 869 AAACGAGCAAAATCTTCATCAATAAAGAAATGTGATGTGTGCAGATCAACAGAGT 928
Qy 481 CCTCTCTTACCTTCTGAAATCCAGATGACAGAGACAGATTCCTCAACACAGAC 540
Db 929 CCGTATTTCTTACCTTCAAAATCCGCACTCAGAGAGACAGATTTCTTCAATACAGAC 988
Qy 541 AAGGAAATGCTGTGTTGAAAGTGTGATTTGATGAGAGAACTTTGGGCTTATCAGC 600
Db 989 AAGGAAATGATAGTTTGAAGAAATGTGATTTGATGATGAGATGACTTGGGCTTATCAGT 1048
Qy 601 TGGTTGCCATCCACCCG 619
Db 1049 TTTTCATTCAGCAAGTCTG 1067

RESULT 3

US-09-777-710A-2
; Sequence 2, Application US/09777710A
; Patent No. US20020058305A1
; GENERAL INFORMATION:
; APPLICANT: OKINO, No. US20020058305A1omu et al.
; TITLE OF INVENTION: CERAMIDASE GENE
; FILE REFERENCE: 1422-0458P
; CURRENT APPLICATION NUMBER: US/09/777,710A
; CURRENT FILING DATE: 2001-02-07
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 2
; LENGTH: 1941
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-777-710A-2

Query Match 5.9%; Score 121.2; DB 10; Length 1941;
Best Local Similarity 49.7%; Pred. No. 8.6e-29;
Matches 340; Conservative 0; Mismatches 338; Indels 6; Gaps 1;

Qy 9 CTACTACATTTGGGCTTGGAGAGCGGATTCACAGACAGTGTCAATATGAT 68
Db 12 CTACCGCTTGGGCTTGGGAGAGCGGATTCACAGAGCGGCGCAAGTCCGATGAT 71
Qy 69 GGGCTATGGCAAAATGCGCAGATGACAGGGGTCTCTACACAGGCTTTCAGCGGTGC 128
Db 72 GGGCTATGGCAAAATGCGCAGATGACAGGGGTCTCTACACAGGCTTTCAGCGGTGC 131
Qy 129 TTTTATCTTGGGATTCAGATGGGTCAAAATGCAATGCAATTTGTGAGCGTGAATATG 188
Db 132 CTTCGTATGAGAGAACGGCGCAGGAGCGTGGCTGTCTACGTCAACACGACCTGGG 191
|||||

OY		189	TATGTTTTCCAAAGCAGCTAGSGTTGGAGSTGCTTGAAGACACTAGAAGATTAATATGTGC	248
Db		192	GATGACCTTCACAGGCCGTGACACCTAAGGTTCTTGCGCCGCTCAACGGCAAGTACC	251
OY		249	TCTGATATGAAGAGACAATGTATTCTGTGATGCTCATTCACACACA.TTGCGGCCACAGG	308
Db		252	TGTCTACACAGCAAGAACAGTATGTGCTGCGCGCACCCACACCACATTCGCGTCCGGGCG	311
OY		309	GTTTTTCCAAATATACACTTATATATCTGCCACGAGGAGATTTCAG.CAACCGSACTTTCA	368
Db		312	CTTTCGCCACATACGAGATGTACAACTGTGCGGTGCTTCCAGAAAAGACTTTCAA	371
OY		369	GTACATAGTCTCTGTGGATCATGAAGACATTGATATATAGTCA.CACACAAATCTTTAAACGAG	428
Db		372	CGCCATCTCGACGGCAGCTGTGCGCTTCATCAGAGGGGGCCCAAGGCTAGGTGACGCCCG	431
OY		429	CAAATCTTTATCAACAAAGGAATGTGCTATATGTGACGATCAACCGAAGCCCTTC	488
Db		432	CCGCGCTGTTTCACGGCAGCGGACCTGGCCAAGCCACCGC.AMCGTTGCTGCTGTC	491
OY		489	TTACCTTGATCCACAGTCAGAGAGACAGATTTCTTCAACACAGACAAAGAAAT	548
Db		492	GCACCTGAAGAATC-----CGGACATGCGCGGCTTAGAGATGEGCATGACCGGAGAT	545
OY		549	GCTGGTCTTGAACATGCTGATTTGAATGSAAGACTTGGGCTTATACAGCTGCTTGC	608
Db		546	GAGCGTGTCACTTCGTCTGACGCCAACGGCAGCGTGGCGGCGGATGATGTTGGTTCCC	605
OY		609	CATCCACCCCGTAGCATGAACAAATAGCAACCACTTTGTGAATAT.GACAAATAATGGCTA	668
Db		606	GGTGACACGACACTGATGACCAACGCCCAATCATCTGATCTCCCCGACACAACAAAGGCTA	665
OY		669	TGCGGCTTACCTTTTGAACAAGA	692
Db		666	CGCTCTCTATCACTGGGAGACAGA	689

```

RESULT 4
US-09-777-710A-16
: Sequence 16, Application US/09777710A
: Patent No. US20020058305A1
: GENERAL INFORMATION:
: APPLICANT: OKINO, No. US20020058305A1omv et al.
: TITLE OF INVENTION: CERAMIDASE GENE
: FILE REFERENCE: 1422-0458P
: CURRENT APPLICATION NUMBER: US/09/777,710A
: CURRENT FILING DATE: 2001-02-07
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 16
: LENGTH: 2013
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-777-710A-16

```

Query Match	5.9%	Score 121.2	DB 10	Length 2013
Best Local Similarity	49.7%	Pred. No. 8.8e-29		
Matches 340	Conservative	0	Mismatches 338	Indels 6
				Gaps 1

QY 9 CTACTACATNTGGCGTTGGGAGAGCGGATTTGCACAGACAAAGTGTGAGATATCAATCAATTTGAT 68
 Db 84 CTACCGCTTGGCGCTCGGGCAGAGCGGACATCCACGGCCAGCCGCCAAGTCGGCATGAT 143
 QY 69 GGGCTATGGCAAAAATATGGCCAGAAATGCACGGGGTCTCTCCACAGGCTTTACGCCGTG 128
 Db 144 GGGTACTCTCTCTCGAAACAGAAACCGCGCGGCATCCACATGCGCCAGTTGGGCGCGTGC 203
 QY 129 TTTTATCTTGGCGGATTCACAGATGGGTCAAAATGCAATGGCATTTTATAGGCTGGAACATTATG 188
 Db 204 CTTCGTGATCGAGGAAGCGGCCAGCGGACGTGGCTGCTCTACGTAAACCGACCTGGG 263
 QY 189 TATGATTTCCCAACGACTGAGGTGTGAGGTCTCTGAAAGAACTAGAGATAAATATATGCTC 248

Dd	264	GATGACCTTCCAGGCGCGTCACCTGAAAGTCCCTGGCCCGGCTCAAGGAGAAAGTACCCCG	3233
Oy	249	TCGTGATGAGAGACAATGTTATCTCGAGTGCATTCACACACACTCTG6CCGACGAG	3089
Dd	324	TGTTCTAGACGAAACAACGTGATGCTCCGCCGCCACCCACACCACTCCGGTCCGGCCGG	3833
Oy	309	GTTTTTCCAAATATACACTCTATATATCTCGCCAGCGAGGATTTCCAGCAACCGGACCTTTCA	3688
Dd	384	CTTCTCCCACTACGCCATGTACAACTCTCGGTCGTCGCGCTTCCAGGAAAAAGACTTCAA	4433
Oy	369	GTAACAATGTCYTGCGGATACGAAGCATTTGATATAGCTCACACAAATCTTTAAACAGG	4288
Dd	444	CGCATAGTGGACGCGATGTCGCGCTCCATCCAGCGGGGCCAGGCGCAGTGTGCAGCCCGG	5033
Oy	429	CAAAATCTTATCAACAAGGAAATGTCTATGTGCAATCAACAACGAAACCCCTCTC	4888
Dd	504	CCGCGCTGTTTACGGCAGCGCGGAGCTGCGCACGCCACCGCGCAACCGTTGCTGCTGTC	5633
Oy	489	TTACCTCTGATCCACAGCTCAGAGAGAGCAAGATTTCTTCAACACAGCAAGAAGAAAT	5488
Dd	564	GCACTGAAGAATC-----CGAGCATCGCCGGTACAGAGATGGCAATCAGCCGCGAGAT	6177
Oy	549	GCATGCTTTAAACTCGTGGAATTTGAATGGAATAAATCTGGGTCTTATCACCTGGTTGC	6088
Dd	618	GAGCGTCTAGCTTCTGTCGACGCCAACGCGCAGCTGCGCGCGCGATCACTGATTTGGTTCC	6777
Oy	609	CATCACACCCCTGAGCATGAAACAATAGCAACCACTTTTGATATAGTACCAATATGGGCTA	6688
Dd	678	GGTGCACAGCACTGATGACCAACGCGCAATACACTGATTTCCCGGACCAACAAGGCTA	7377
Oy	669	TGCGGCTTACCTTTTGGAGCAGA	692
Dd	738	CGCGTCTATCACTGGAGCACGA	761

```

, RESULT 5
, US-09-878-574-9145
, Sequence 9145, Application US/09878574
, Patent No. US20020110548A1
, GENERAL INFORMATION:
, APPLICANT: Byrum, Joseph R.
, APPLICANT: La Rosa, Thomas J.
, APPLICANT: Thompson, Michael D.
, TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
, TITLE OF INVENTION: Plants
, FILE REFERENCE: 38-21(15401)B
, CURRENT APPLICATION NUMBER: US/09/878,574
, CURRENT FILING DATE: 2001-12-21
, PRIOR APPLICATION NUMBER: 09/333,535
, PRIOR FILING DATE: 1999-06-14
, NUMBER OF SEQ ID NOS: 15775
, SEQ ID NO 9145
, LENGTH: 272
, TYPE: DNA
, ORGANISM: Glycine max
, OTHER INFORMATION: Clone ID: 70110213BH1
, US-09-878-574-9145

```

Query Match	3.5%	Score 72.2	DB 10	Length 272
Local Similarity	60.0%	Pred. No. 4.1e-13		
Matches 141:	Conservative	0	Mismatches 88	Indels 6
			Gaps	1:
QY	1291	GTGGGTCCTTGGCCATGACTGCTATCCCTGGGAAATTACACCATGTCGGGACGANGA	1350	
DB	15	GTGGGCGAGCTTGTTATTCTCAGTGTACTCTGGAGATTTCACACACGCGTGGGAGCGT	74	
QY	1351	TTTCGTGAGSCAATTAAGAAATTGGCACTTATGGAGTGAAGATATGACC-----	1404	
DB	75	CTTCGTGATGCAGTGAAGACAGTCAAGTGTGTAACAAAGGCTTTGGTAGCAACATTCAT	134	
QY	1405	GTTCGTATCCGAGGCTTACAGCAATGTTATACACATTACATTACCACATATGAGAAATAC	1464	
DB	135	GTTGTATTAGGGGGTTGACTTAATACCTATTTCACAGTATGGCACTTACATTACGAGAGATAC	194	

QY 1465 CAGCTCAGCGGTACGAGCAGCATCTACATCTATGACACACCCCTGCTG 1519
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
Db 195 CAGGTGACAGATATGAGGGTGTCTCCACAGTATGTCACACACTGAGTG 249

RESULT 6

US-09-910-943-91
Sequence 91, Application US/09910943
Patent No. US20020081610A1

GENERAL INFORMATION:

APPLICANT: Hemmati-Bryanlou, Ali
APPLICANT: Altman, Curtis
TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
FILE REFERENCE: 7529/1G148US1
CURRENT APPLICATION NUMBER: US/09/910,943
CURRENT FILING DATE: 2001-07-23
NUMBER OF SEQ ID NOS: 742
SOFTWARE: PatentIn version 3.1
SEQ ID NO 91
LENGTH: 713
TYPE: DNA
ORGANISM: Xenopus laevis
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(713)
OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-91

Query Match 1.8%; Score 39.8; DB 10; Length 713;
Best Local Similarity 60.7%; Pred. No. 0.051;
Matches 65; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 357 CCGACCTTTCAGTACATGCTCTGCGATCATGAAGACATGATAGCTCACAAA 416
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
Db 115 CCGCGCATTTCTACACAAATGTCACCTGAGCCATCAAGAGAGTATGACTTAAAAA 174

QY 417 TCTTAACCAAGCAAAATCTTTATCAACAAGGAATGTTGCTAATG 463
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
Db 175 CCTGAAGAAGGAGACATAATATCTCAATGAGAGACTGTAGTG 221

RESULT 7

US-09-954-456-1057/c
Sequence 1057, Application US/09954456
Patent No. US20020115057A1

GENERAL INFORMATION:

APPLICANT: Young, Paul

TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand

FILE REFERENCE: 689290-76

CURRENT APPLICATION NUMBER: US/09/954,456

CURRENT FILING DATE: 2001-09-18

PRIOR APPLICATION NUMBER: US/60/233,617

PRIOR FILING DATE: 2000-09-18

PRIOR APPLICATION NUMBER: US/60/234,052

PRIOR FILING DATE: 2000-09-20

PRIOR APPLICATION NUMBER: US/60/234,923

PRIOR FILING DATE: 2000-09-25

PRIOR APPLICATION NUMBER: US/60/235,134

PRIOR FILING DATE: 2000-09-25

PRIOR APPLICATION NUMBER: US/60/235,637

PRIOR FILING DATE: 2000-09-26

PRIOR APPLICATION NUMBER: US/60/235,638

PRIOR FILING DATE: 2000-09-26

PRIOR APPLICATION NUMBER: US/60/235,711

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: US/60/235,720

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: US/60/235,840

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: US/60/235,863

PRIOR FILING DATE: 2000-09-27

NUMBER OF SEQ ID NOS: 2276
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1057

LENGTH: 462

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

OTHER INFORMATION: n=a,t,g or c

US-09-954-456-1057

Query Match 1.8%; Score 37.4; DB 10; Length 462;
Best Local Similarity 57.8%; Pred. No. 0.23;
Matches 85; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

QY 1602 ATTCTCAAAAATCTAATAGCTTCACTTATTCCTAATATTTGGGATAGACCAATTTGG 1661
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
Db 319 ATTTTCAAAAATTTTATATACATGACATATTTACATGGCTTAAGCAAGACTTGG 260

QY 1662 CAACATTTTGGGATGTCCTGCAGCCAGCAAAACCTGAATACAGAGTGGAGAACTGGT 1721
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
Db 259 CCACA-AAATACCTATTGTGCTGTAATGAATACATGATTAAGCAAGCTGTGTAGC 201

QY 1722 TGAATTTATTTGTAGCGGCTTAACC 1748
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
Db 200 TGAATTTATGAGGAATCCCAATCTC 174

RESULT 8

US-09-878-574-8039
Sequence 8039, Application US/09878574
Patent No. US20020110548A1

GENERAL INFORMATION:

APPLICANT: Byrum, Joseph R.

APPLICANT: La Rose, Thomas J.

APPLICANT: Thompson, Michael D.

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

FILE REFERENCE: 38-21(15401)B

CURRENT APPLICATION NUMBER: US/09/878,574

CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: 09/333,535

PRIOR FILING DATE: 1999-06-14

NUMBER OF SEQ ID NOS: 15775

SEQ ID NO 8039

LENGTH: 170

TYPE: DNA

ORGANISM: Glycine max

OTHER INFORMATION: Clone ID: 701100675H1

US-09-878-574-8039

Query Match 1.8%; Score 36; DB 10; Length 170;
Best Local Similarity 60.0%; Pred. No. 0.32;
Matches 60; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1173 AAACCATCTGAAGAGATTGAGTGCAGAAACCAACCAATCTCTTCACAGTGG 1232
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
Db 49 AACGCCAGCAGCAAGAACAGTAGACGTGCACCAACCCCAATTTCTGATGACTGG 108

QY 1233 AGACGTGAGATACATCTTGGCAACAGATATTGTT 1272
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
Db 109 CGAATGAAGCTACATATGATGGGCGCTTCATACTT 148

RESULT 9

US-10-108-605-102
Sequence 102, Application US/10108605
Patent No. US20020160934A1

GENERAL INFORMATION:

APPLICANT: Broadus, Julie

APPLICANT: Stam, Lynn

APPLICANT: Bachmann, Jane

APPLICANT: Kamdar, Kim

;; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
;; FILE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
;; FILE REFERENCE: 31133B
;; CURRENT APPLICATION NUMBER: US/10/108,605
;; PRIOR APPLICATION NUMBER: US 09/761,142
;; PRIOR FILING DATE: 2001-01-16
;; PRIOR APPLICATION NUMBER: US 60/176,418
;; PRIOR FILING DATE: 2000-01-14
;; NUMBER OF SEQ ID NOS: 361
;; SOFTWARE: Patent In Ver. 2.1
;; SEQ ID NO 102
;; LENGTH: 14155
;; TYPE: DNA
;; ORGANISM: Drosophila melanogaster
US-10-108-605-102

Query Match 1.7%; Score 34; DB 9; Length 14155;
Best Local Similarity 51.3%; Pred. No. 40;
Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 936 GCTGTATGCTCTCTCCCTCCAGGAGTGACCGGCCCACTGCTTGAGCTCACCACTGGCT 995
DB 10484 GCACTTGGGCGAGAAATCGAACTTGAAGGCCCACTGAGGCTGCTCCCACTTGGC 10543
QY 996 GAACATGACAGATGTGAGCGTCCAGCTCAATGCCACACACAGTAGAGAGCTGTAAAC 1055
DB 10544 CAACAGCATCAAGTGGCGCTCAATTTCAAGCCAAAGCAGTCTCTGAGACTGAAGACACC 10603
QY 1056 TGCCCTGGGCTACAGTTTGGCCGACGACCAATT 1089
DB 10604 GGAAGAGCAAGTTGCTAGCCACTCGACCAAT 10637

RESULT 10

US-09-822-849A-310/C
; Sequence 310, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Mong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 310
; LENGTH: 1898
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-310

Query Match 1.6%; Score 33.6; DB 10; Length 1898;
Best Local Similarity 57.7%; Pred. No. 12;
Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1792 GTGAGAAATCGAGACTCTGTAGCTGACTGCGAGATATGATATACGATGCTCTGG 1851
DB 1705 GGGATGAAGAGATGACCTGTGGTGGCGTGACAGCAGCAGTGTGAGATTGATCTCAGG 1646
QY 1852 GAGACGAGTTTATTGGCACAAGGAATACTGGGTCTGACCAA 1895
DB 1645 GGGACAGCTGACACAGGCACTGGGAATCTTGGGCAAGGACAA 1602

RESULT 11

US-10-198-846-2496
; Sequence 2496, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2496
; LENGTH: 415
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 5
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-2496

Query Match 1.6%; Score 33.4; DB 9; Length 415;
Best Local Similarity 65.3%; Pred. No. 4.6;
Matches 49; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 709 CTGCTGACAGAGGACGTTGTAGCAGCTTTGCTTGCATCAATCTCGAGAGCTGTCA 768
DB 58 CCGCCCGGCGAGTCAATCTGAGCTGCGCTTCATCATCATCTGCAAGACTCATCG 117
QY 769 CCCAATCTCTTGGC 783
DB 118 CCAAGCATCTCTGCG 132

RESULT 12

US-10-184-644-518
; Sequence 518, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 518
; LENGTH: 1137
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-518

Db 473 KRISOIKSKKRCSSGSEDEYRSRFSSECFMDLVCEKRCBEGTIVDC:SNOKIVRIPSHPE 532
OY 645 TGTGAATGATGCAATATGAGGCTATGCGCTTACTTTTGAGAGAGAAAGAACAAAGG 704
Db 533 YTTDLRLNDNEVSVLEATGIFRKLPLNLR-INLSNNKIKEYREGAFDGAASVDELMLTGN 591
OY 705 CTATCTGCTGAGACAGGAGCGCTTTGTAGACAGGCTTTGCTTCAATCAATCTCGAGAGCT 764
Db 592 QLETVHGVFRGLSLKTLMLRSLNLSVSNDFAGLSSVRLSLYDRIRITTPGATTT 651
OY 765 GTCACCCCAACATTCTTGCGCCCGCATTTGTCAACACAGGAGAGCTTTGTCAACAGCA 824
Db 652 LVSLSTINLNSPNCNCHLMLGKWLKRRIRIVSGNPRCQRPFLKEIPIDVAIODEFTC 711
OY 825 GAGACCTGTCCCAACGCTGGGCGCTAGCATGTGCATGCGCCAGCGAGCTGTGACAAAGACAT 884
Db 712 DNESSCQLSPRCBECTCMEYVRCNKGRLALPRMPKDVETLLEGNHLRAVPREL 771
OY 885 GTTGAAGACACACATTATAGACGAGATCATCTATCAAGAGCAAGAGAGCTGTATGC 944
Db 772 SALRHLTLDLSNNSISMLTYTFSSNMHSLTLLISYRLRCIPVHAENGRLSLRVLTFLH 831
OY 945 CTGCGCTCCAGAGAGGTGACCGGCCAGTCTTGACAGTCAACAGTGGTGAACATGAC 1004
Db 832 GNDISSYVEGSEFNDLTSLSHLALGTNPJLHCDLSRLMSEWYKAGYKEPGIARCSPEBMA 891
OY 1005 AGATGTGAGCGTCCAGCTCAATGACACACACAGTGAAGAGTGAACCTGCGCCTGG 1064
Db 892 DRLLITPTTHRQCGPYDINIVAKCNACLSPPCKNNCTQDPVELYRCACPYSYKGD 951
OY 1065 CTACAGTTTTCGCCGAGGACACAAATGA 1091
Db 952 CTVPIINICIOINPCQHGCTCHLSDSHKD 978

RESULT 15
US-10-184-634-290
; Sequence 290, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jlan
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 290
; LENGTH: 1523
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-290

Query Match 1.68, Score 33, DB 9, Length 1523;
Best Local Similarity 9.1%; Pred. No. 16;
Matches 57, Conservative 163, Mismatches 406, Indels 1, Gaps 1;

OY 465 GCAGATCAAGCGAGAGCCCTCTTACCTCTGATCAACAGATCAAGAGAGAGAGAGTA 524
Db 353 GLKSLTSLVLYGNKITEIAKLPDGLVSLQLLLNANKINCLRVNTFODLONLNLSLYD 412
OY 525 TTCTTCAACACAGACAGAAATGCTGCTTGAACCTGCTGATTTGAATGAGAGAGA 584

Db 413 NKLOITISKGLPAPLQSIQTLHLAQNPFVCDCHLKLADYLDNPLETSGARCSSPRLAN 472
OY 565 CTGGGCTTTATGAGCTGTTGCGATCCACCCTGTGAGCATGAAACATATACCAACACTT 644
Db 473 KRISOIKSKKRCSSGSEDEYRSRFSSECFMDLVCEKRCBEGTIVDC:SNOKIVRIPSHPE 532
OY 705 TGTGAATGATGCAATATGAGGCTATGCGGCTTACTTTTGAGAGAGAAAGAACAAAGG 704
Db 533 YTTDLRLNDNEVSVLEATGIFRKLPLNLR-INLSNNKIKEYREGAFDGAASVDELMLTGN 591
OY 645 CTATCTGCTGAGACAGGAGCGCTTTGTAGACAGGCTTTGCTTCAATCAATCTCGAGAGCT 764
Db 592 QLETVHGVFRGLSLKTLMLRSLNLSVSNDFAGLSSVRLSLYDRIRITTPGATTT 651
OY 765 GTCACCCCAACATTCTTGCGCCCGCATTTGTCAACACAGGAGAGCTTTGTCAACAGCA 824
Db 652 LVSLSTINLNSPNCNCHLMLGKWLKRRIRIVSGNPRCQRPFLKEIPIDVAIODEFTC 711
OY 825 GAGACCTGTCCCAACGCTGGGCGCTAGCATGTGCATGCGCCAGCGAGCTGTGACAAAGACAT 884
Db 712 DNESSCQLSPRCBECTCMEYVRCNKGRLALPRMPKDVETLLEGNHLRAVPREL 771
OY 885 GTTGAAGACACACATTATAGACGAGATCATCTATCAAGAGCAAGAGAGCTGTATGC 944
Db 772 SALRHLTLDLSNNSISMLTYTFSSNMHSLTLLISYRLRCIPVHAENGRLSLRVLTFLH 831
OY 945 CTGCGCTCCAGAGAGGTGACCGGCCAGTCTTGACAGTCAACAGTGGTGAACATGAC 1004
Db 832 GNDISSYVEGSEFNDLTSLSHLALGTNPJLHCDLSRLMSEWYKAGYKEPGIARCSPEBMA 891
OY 1005 AGATGTGAGCGTCCAGCTCAATGACACACACAGTGAAGAGTGAACCTGCGCCTGG 1064
Db 892 DRLLITPTTHRQCGPYDINIVAKCNACLSPPCKNNCTQDPVELYRCACPYSYKGD 951
OY 1065 CTACAGTTTTCGCCGAGGACACAAATGA 1091
Db 952 CTVPIINICIOINPCQHGCTCHLSDSHKD 978

Search completed: July 5, 2003, 12:55:00
Job time : 344 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 07:07:38 ; Search time 101 Seconds
(Without alignments)
6221.591 Million cell updates/sec

Title: US-09-937-521-15
Perfect score: 2049
Sequence: 1 ttcaagtgactactactg.....ttgaagttgactactag 2049

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
Issued_Patents_NA:*
1: /cgn2.6/prodata/1/lna/5A.COMB.seq:*
2: /cgn2.6/prodata/1/lna/5B.COMB.seq:*
3: /cgn2.6/prodata/1/lna/6A.COMB.seq:*
4: /cgn2.6/prodata/1/lna/6B.COMB.seq:*
5: /cgn2.6/prodata/1/lna/PCTUS.COMB.seq:*
6: /cgn2.6/prodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	121.2	5.9	1941	4	US-09-328-501-2
2	121.2	5.9	2013	4	US-09-328-501-16
3	80	3.9	4411529	4	US-09-103-840A-1
4	74.2	3.6	4403765	4	US-09-103-840A-2
5	37.2	1.8	7218	1	US-08-232-463-14
6	36.4	1.8	1716	3	US-08-656-034-9
7	36.4	1.8	2160	3	US-08-656-034-1
8	33.4	1.6	1608	4	US-09-134-001C-2407
9	33.2	1.6	1882	1	US-08-257-073-12
10	33.2	1.6	1884	1	US-08-257-073-8
11	33	1.6	1554	4	US-09-134-001C-2607
12	32.6	1.6	2518	4	US-09-011-745-1
13	32.6	1.6	5865	4	US-09-011-745-8
14	32.4	1.6	2172	4	US-08-030-410-2
15	31.8	1.6	1021	4	US-09-465-558-61
16	31.6	1.5	500	3	US-08-545-809A-37
17	31.6	1.5	531	4	US-08-642-274D-24
18	31.4	1.5	531	4	US-08-952-014C-24
19	31.4	1.5	2869	4	US-08-961-527-130
20	31.4	1.5	289	4	US-09-007-005-17
21	31.2	1.5	289	4	US-09-244-796-17
22	31.2	1.5	1140	2	US-08-896-410-1
23	31	1.5	1227	4	US-08-961-527-250
24	31	1.5	1925	4	US-08-709-731A-1
25	31	1.5	3814	4	US-09-302-812-5
26	31	1.5	3814	4	US-09-511-477-5
27	31	1.5	3814	4	US-09-511-507-5

C	28	31	1.5	26700	1	US-08-472-217-1	Sequence 1, Appl
C	29	31	1.5	26700	2	US-08-488-199-5	Sequence 5, Appl
C	30	31	1.5	26700	3	US-08-760-534A-1	Sequence 1, Appl
C	31	30.8	1.5	2184	1	US-07-815-333A-1	Sequence 1, Appl
C	32	30.8	1.5	2288	1	US-08-230-937B-4	Sequence 4, Appl
C	33	30.8	1.5	2289	1	US-07-838-410-2	Sequence 2, Appl
C	34	30.8	1.5	4086	4	US-09-221-017B-363	Sequence 363, App
C	35	30.2	1.5	773	4	US-08-998-416-385	Sequence 385, App
C	36	30.2	1.5	804	2	US-08-767-026-12	Sequence 12, Appl
C	37	30.2	1.5	2277	1	US-08-676-967-2	Sequence 2, Appl
C	38	30.2	1.5	2277	1	US-08-676-974-2	Sequence 2, Appl
C	39	30.2	1.5	2277	2	US-09-098-487-2	Sequence 12, Appl
C	40	30.2	1.5	3209	1	US-07-803-633A-12	Sequence 79, Appl
C	41	30.2	1.5	9144	4	US-08-556-978B-79	Sequence 3, Appl
C	42	30.2	1.5	36651	4	US-09-738-894A-3	Sequence 1576, Ap
C	43	30	1.5	420	4	US-09-134-001C-1576	Sequence 4, Appl
C	44	30	1.5	2091	1	US-08-765-081-4	Sequence 4, Appl
C	45	30	1.5	2091	3	US-09-098-082-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1									
US-09-328-501-2									
Sequence 2, Application US/09328501A									
Patent No. 6258581									
GENERAL INFORMATION:									
APPLICANT: ITO, Makoto									
APPLICANT: OKINO, No. 6258581omu									
TITLE OF INVENTION: Ceramidase Gene									
FILE REFERENCE: 1422-0377P									
CURRENT APPLICATION NUMBER: US/09/328, 501A									
CURRENT FILING DATE: 1999-06-09									
EARLIER APPLICATION NUMBER: 10-234769 JAPAN									
EARLIER FILING DATE: 1998-08-20									
NUMBER OF SEQ ID NOS: 18									
SOFTWARE: PatentIn Ver. 2.1									
SEQ ID NO 2									
LENGTH: 1941									
TYPE: DNA									
ORGANISM: Pseudomonas aeruginosa									
FEATURE:									
OTHER INFORMATION: any n or xaa = Unknown									
US-09-328-501-2									
Query Match									
Best Local Similarity 49.7%; Pred. No. 1.8e-30;									
Matches 340; Conservative 0; Mismatches 338; Indels 6; Gaps 1;									
QY	9	CTACTACATTCGCGCTTGGGAGAGCGGATTCGACAGGCAAGTGCATATTCATTGAT	68						
DB	12	CTACCGCTTGGCGCTGGGCAAGGCGACATCCGCGAAGCCCGCAAGTCGGCATGAT	71						
QY	69	GGGCTATGCGAAATATGCGAATGACGCGGCTCTCAACGAGCGCTTCAGCCGTGC	128						
DB	72	GGGTACTCTCTCCCTCGAAGAGAGACCGCGCATCATGCGCCAGTGGCGGCTGC	131						
QY	129	TTTATCTTGGCGGATCCAGATGGGTCAATGCAATGCAATTTGTGAGCGTGAATG	188						
DB	132	CTTGATGATGAGAGACCGCGACGCGACGCGCTGCTGCTACCTCAACCCGACGTGG	191						
QY	189	TATGATTTCCCAAGCACTGAGGTGGAGGCTCTGAAGACATAGAGTAAATATGCTC	248						
DB	192	GATGACCTTCCAGGCGCTGCACTGGAAGGCTCGCCGCTCAAGCGAAGTACCCCG	251						
QY	249	TCTGTATGGAAGACATATGTTATCTGAGAGCCATTCACACACACTCTGGCCAGAG	308						
DB	252	TGCTTAGAGAGAGAACACATGATGCTGCGCCACCCACACACCTCGGTGCGGCGG	311						
QY	309	GTTTTCATATATACACTATATATAGTCCGACGAGGATTCAGCAACCGGACCTTTCA	368						
DB	312	CTTCTCCACTACGCGATGTAACAACCTGTGCTGCTGCTTCAGGAAAGACCTTCAA	371						

OY 369 GTACCTAGCTCCTGGGATCATGAAGACTTGATATAGCTCAACAACAAATTCTTAATAACAG 428
Db 372 CGCCATCTCGAAGGCATGTGTCCTCCATCGAGGGGGGCCAGGCAGGTTGACGGCCGG 431
OY 429 CAAAATCTTATATCAACAAGAAGAAATGTTGCTAATGTGCAGATCAACCGAAGCCCTCTC 488
Db 432 CGGCTGTTCATAGGACGGGAGGAGACTGGCGAACCCGACCCGCMACCCTTCCGTGTC 491
OY 489 TTACCTTGAATCCACAGTCAGAGAGAGCAAGGTAATCTTCAACACAGACAGGAAT 548
Db 492 GCACCTGAAGATC-----CGACATCGCCGGCTACGAGATGGCATGACGCCAGAT 545
OY 549 GCTGGTCTTGAACACTGCTGGATTTAAATGAGAAAGACTTGGGCTTATACGTGGTTTC 608
Db 546 GAGCTGTCTCAGCTTCGTGACGCCAACGGCAGGCTGGCGGGCGGATCAAGTTGTTCCC 605
OY 609 CATCCAGCCCCGAGNACATGAACAAATAGCAACCACTTGGATATGACAAATATGGGCTA 668
Db 606 GGTCGACAGCACCTCGATGATCCCAAGCCCAATACCTGATCTCCCGGACAAAGGGCTA 665
OY 669 TGGCGCTTACCTTTTGGACAGA 692
Db 666 CGCCTCTATCACTGGAGACAGA 689

```

2
US-09-328-501-16
: Sequence 16, Application US/09328501A
: Patent No. 6258581
: GENERAL INFORMATION:
: APPLICANT: OKINO, NO. 6258581ommu
: APPLICANT: ITO, MAKOTO
: TITLE OF INVENTION: Ceramidase Gene
: FILE REFERENCE: 1422-0377P
: CURRENT APPLICATION NUMBER: US/09/328,501A
: CURRENT FILING DATE: 1999-06-09
: EARLIER APPLICATION NUMBER: 10-234769 JAPAN
: EARLIER FILING DATE: 1998-08-20
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 16
: LENGTH: 2013
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: FEATURE:
: OTHER INFORMATION: any n or xaa = Unknown
US-09-328-501-16

```

	Query Match	5.9%	Score 121.2	DB 4	Length 2013	
	Best Local Similarity	49.7%	Pred. No. 1,8e-30			
Matches	340	Conservative	0	Mismatches	338	Indels 6; Gaps 1
QY	9	CTACTACATTGGCGTGGGAGACGGGATTCACACAGACAAAGTGTGAGATATCAATTTGAT	68			
DB	84	CTACCGCTTGGCGCTGGGCAAGCGCGAATCAACGGCAAGCCCGGAAGTGGGATGAT	143			
QY	69	GGCGTATGGCAAAAATGGCCAGAAATGCACGGGGTCTCCACACAGCGTTCAGCCGTGC	128			
DB	144	GGGTACTCTCCCTCGAACAAGAACCGCGCGCATCCACATAGCCCAAGTGGCGCGTGC	203			
QY	129	TTTATATCTTGGCGGATCCAGATGGGTCAAAATGGAATGGCATTTGTGAGCGTGGAACTATG	188			
DB	204	CTTGTGATTCAGGAAGCGGCGACGGGACGTGCGCTGTGATCAAGTCAACACCGACCTGGT	263			
QY	189	TATGATTTCCCAACGACTGAGGTTGGAGTCTCTGAAGAGACTATGAGTAATATATGCTC	248			
DB	264	GATGACCTTCAGGCGCGTGCACCTGAAGGTCTCTGCGCGGCTCAAGGGAAGTACCCGG	323			
QY	249	TCTGTATCGAAGACACATGTTATTCCTGATGTCCTTACACACACTCTGGCCCAAGAG	308			
DB	324	TGTCATACAGAGAACAACGTATGCTCTGCCGCCACCCACACCCACTCCGCTCGGCGCGG	383			

OY GTTTTTCAAATATACACTTATATCTGCGCAGCGAGGGATTCAGCAACCGACTTTC 368
 Db 384 CTCTCCCACTAGCGCATGTACAACTGTGCTGCTGCGCTCCAGGAAAAAGACTTCAA 443
 OY 369 GTACATAGTCTCTGGGATCATGAGAGCAATTGATATAGCTCACAAATCTTAAACGAG 428
 Db 444 CGGCATCGTACAGCGGATCTCGCTCCATCATGAGCGGGCCACGAGCCAGTTGACGCCG 503
 OY 429 CAAAATCTTTATCAACAAGAAATCTTGCTAATGTGACATCAACCGAAGCCCTCTCC 488
 Db 504 CGCGCTGTTCTACGGGAGCGGAGCTCGGACCTCGCAACCCAGCCGCTTGGCTGCTC 563
 OY 489 TTACCTTTGAAATCCACAGTACAGAGAGCAAGGATTTCTTCAACAACAGCAAGAAAT 548
 Db 564 GCACCTGGAAGATC-----CGGACATGTGCGCGCTACGAGGATGGCATCGACCCGAGAT 617
 OY 549 GCTGTCCTTGAAACTGTGTGATTTGTAATGGAAGAACTTGGTCTTATCACTGGTTGC 608
 Db 618 GAGCGTGTACAGTGTGTGACGCCAACGGGAGAGTGGCGGCGGATCGAGTTGCC 677
 OY 609 CATCCACCCCGTGSAGCATGAAACAATAGCAACCACTTTGTGTAATGTACATATGGGCTA 668
 Db 678 GGGGACACACACTGTGATACCAACGCCAATTCACGTATCTCCCGGAGCAACAAGGGCTA 737
 OY 669 TCGGCTTACCTTTTGACAGAGA 692
 Db 738 CGCTCTCTATCACTGGGAGACAGCA 761

```

RESULT 3
US-09-103-840A-1/c
: Sequence 1, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103.840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 1
: LENGTH: 4411529
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

	Query Match	Best Local Similarity	3.98;	Score 80;	DB 4;	Length 4411529;
	Matches	380;	Conservative	0;	Mismatches 430;	Indels 12; Gaps 3;
OY	25	GGGAGAGCGGATGTCACAGACAAAGTGTACATATCAATTGATGGCTATGCGAAAAAT	84			
Db	769577	GGCATGCGCCGACATCATCTGTGAGGCGGGGACCTCGGCATGCTCGCGTAGCGAGATGCC	769518			
OY	85	GGCGAGAAATGACAGGGGGTCTCCATCACAGGCGTGTCAAGCCGTGTTTATCTTGGCGGA-	143			
Db	769517	GACACAGCGACCGCCGGCATTTACCAAGGACACTTACGTACAGGGCGTTTGTGTCCGGGAC	769458			
OY	144	-----TCCAGATGCGTCAAAATCGAAATGGCATTTGTGACGCGTGAACATATGATGATTTCC	198			
Db	769457	GATTCCCAAGACGGGGGACGCGGGTGTGCTGATGTTCGCGCAACTCCCGCTGCCCATG	769398			
OY	199	CAATGACGTGAGGTTGGAGGTCTCGAAGAGACTAGAGATTAATATTTGGCTCTCTATCGA	258			
Db	769397	CAAAAGCTGAATAGAGAAATGCTGGCGCGGTTTGTATGAGCGAATACCTACTCG	769338			
OY	259	AGAGACAATGTTATCTCGATGCCAATTACACACACTCTGGCCACGAGGGTTTTCCAA	318			

Db 769337 GAGCAGAAACACATGATACAGGGGACCCACACCCCGGGCTT3CGGGTACTGCGGA 769278
QY 319 TATACACTATATACCTCCGAGGAGGATTCAGCAACCGGACCTTTCAGTATATGTC 378
Db 769277 TACCTGCTGTACACCTGACACGAGCGGCTTTCGCGGCAACCTTTCGCGCATGTC 769218
QY 379 TCTGGATCATGAGACATGATATATAGCTCACAAATCTTAAACGAGCAAAATCTT 438
Db 769217 GAGGAAATTTGCAATGGGGAACAATGCCATGCCATGATGGGCCCGCGAGTTTCG 769158
QY 439 ATCAACAAGAAATTTGCTTAATGTGCAATCAACCGCAAGCCCTTCCTTTACCTTCTG 498
Db 769157 CTCTCGACCGCGGCTGTACAGGGCAAGCATCATGATCCCATGAGGCTTGACCGC 769098
QY 499 AATCCAGTCAGAGAGAGAGATTTCTTCAAAACAGACAAAGAAATGCTGTCTG 558
Db 769097 AACCCGCGCGCGCAAGGGCTTCTCCCAACGCGGTGACCCG3CATACACGCTGTC 769038
QY 559 AAACGTGATTTGTAATGGAAGAAAGACTTGGTCTTATCAGCTGTGTTGCCATCACCC 618
Db 769037 ---CGATGACCGGGGCGAGGCCACCGTGGGCTGATTCATTTCTTCCGTACCCAGGT 768981
QY 619 GTAGCATGAACATACCAACCACTTTGTGAATAGTGAATATGAGCTATGCGGCTTAC 678
Db 768980 ACAGCATGAGCATGCAACCAACCTCATCTCCGAGAGCAAAAGGTTTCGCGGCTAT 768921
QY 679 CTTTGTAGCAAGAAAG---AACAAAGCTATCTGCTGAGACGAGGACCGTTGTGCA 735
Db 768920 CACTGGAGACGACCGGTGGGCGCGGATATCTCCGCGCGCATTCCTCATGCGC 768861
QY 736 GGGTTTGCTTCATCAATCTCGAGAGCTGTCAACCAATCTTGGCCGCAATGTGTC 795
Db 768860 GCCTTCCAGAGAGATCCGGGCGACATGAGCCGGAATGTCCAGGGGCCCTTCTCC 768801
QY 796 AACACAGGGGAGTCTGTGTGACAAAGACAGACACCTGTCC 837
Db 768800 GAGGCCCCACACGACCGGGAATTCACACACCCGCTGTACC 768759

RESULT 4
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 3.6%; Score 74.2; DB 4; Length 4403765;
Best Local Similarity 45.8%; Pred. No. 4.8e-12;
Matches 377; Conservative 0; Mismatches 433; Indels 13; Gaps 3;

QY 25 GGGAGAGCGATTCACAGACAGCAAGTTCAGATATCAATTTGATGGCTTGGCAAAAT 84
Db 771538 GGCATGCGCCGACATCACTGCTGAGGGCGGCGACTGCGCATGCTCGGCAAGACTCC 771479

QY 85 GGCCAGATGACGCGGGCTCTCCACCAGGCTGTTACGCCGTATTATCTTGGCGGA- 143
Db 771478 GACCAGGACCGCGCGCATTCACACGACTTATGATGACGGGCTTGTGTTCCGGGAC 771419
QY 144 -----TCCAGATGGGTCAATATCAATGGCATTTGTGAGCGTGAACATATATGATTTCC 198
Db 771418 GATTCCAGAGCGGGGACCGGGGTGCTGCTGATGTGTCGGCAATGCGGCCATG 771359
QY 199 CAACGACTGAGGTGAGGCTCTTAAGAGACTAGAGATTAATATGCTCTGTATCGA 258
Db 771358 CAAACGTAATGAGGAAGTGTGCGCGGCTTGGCCGATTTGATATGCGATACCTACTCG 771299
QY 259 AGAGACATGTTATTCAGAGTCCGATTCACACACACTCTGCGCCAGAGGTTTTCAA 318
Db 771298 GAGCAGAAACACACTGATCAGCGGACCCACACCCAGCGCGCTGCGGCTACTGCGGA 771239
QY 319 TATACACTATATATCTGCCAGGAGGATTCAGCAACCGGACCTTTCAGTATAGTC 378
Db 771238 TACCTGCTGTACACTGACACAGGCGGCTTTCGCGGCAACCTTTGCGGCGATGCTC 771179
QY 379 TCTGGATCATGAGAGCATTTGATATAGCTTACACAAATCTTAAACGAGCAAAATCTT 438
Db 771178 GACGGAATTTGCAATGCGTGAACATGCGCATGCCATGCGCGCCGCGAGTTTCG 771119
QY 439 ATCAACAAGAAATTTGCTTAATGTGCAATCAACGAGGACCCCTCTTACCTTCTG 498
Db 771118 CTCTCGACCGCGAGCTGTACGGCGCAACATCATATGATCCCATATGAGCTTGACCGC 771059
QY 499 AATCCAGTCAGAGAGAGCAAGGATTTCTTCAAAACAGACAAAGAAATGCTGTCTG 558
Db 771058 AACCGCGCGCGCGCAAGGGCTTCTCCGAAAGCGTGTGACCGCGCATACAGCTGTG 770999
QY 559 AAACGTGATTTGTAATGGAAGACTTGGGTCTTATCAGAGTGTGTTGCCATCACCC 618
Db 770998 ---CGTATCGACCGGGCGAGGCGCACCGTGGCGCTGATCTTCTGCTACACAGGT 770942
QY 619 GTGACATGAACATACCAACCACTTTGTGAATAGTGAATATGAGCTATGCGGCTTAC 678
Db 770941 ACAGCATGAGCAATTCGCAACCACTCATCTCCGAGACAAAGGTTTCGCGGCTAT 770882
QY 679 CTTTGTAGCAAGAAAGAAAGAGCTAT---CTCCGTGACAGGAGCGGCTTGTAC 734
Db 770881 CACTGGAGACGACCGTGGGCGCGGATATATCTCCGCGCAAGCCCGCATTCATGCG 770822
QY 735 AGCTTTGCTTCATCAATCTCGAGACGTGTCAACCAATTTTGGCGCGATTTGT 794
Db 770821 CGCTTTCAGAGCAAGATCCGGGCGACATGAGCCCAATGTGACGCGGCCCTTCTCC 770762
QY 795 CAACACAGGGGAGTCTGTGTGCAACGACAGAGGACCTGTCC 837
Db 770761 CGAGGCCCCACACGACCGGGAATTCGACACACCCGCTGTACC 770719

RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PT29pt-fls
US-08-232-463-14

Query Match 1.8%; Score 37.2; DB 1; Length 7218;
Best Local Similarity 5.5%; Pred. No. 0.19;
Matches 21; Conservative 194; Mismatches 167; Indels 0; Gaps 0;

1644 GGATGAGCACCACCAATGGCAACATTTGGGAGTCTTCGACGACCAAAACCTGAAAT 1703
DB RRR 1345
1704 CAGAGTGGAGAGAGTGTGAATATTTGTAGCCGCTAACCAAGAAATTCAGACA 1763
DB RRR 1285
1764 GAACGAGCCCATCAACCTCTCTGCTGAGCAATACGAGAGCTGTGCTGACTG 1823
DB RRR 1225
1824 GCAGATATGTATATGATGCTCTCTGAGAGAGAGCTTTTATGACCAAGAAATACT 1883
DB RRR 1165
1884 GGGTCTGAGCAATGCAATATATCTGCAATATTCAGATATCTGCTGCAATCTA 1943
DB RRR 1105
1944 CAGATAGATATTTTGGACCAATCGGAGAGAGAACTTTCGAACCCGCTGTATACT 2003
DB RRR 1045
2004 AGCATTTGAGGAATTTCTCT 2025
DB 1044 AGCCAAGCTCGGAATTAATTTCT 1023

RESULT 6

US-08-656-034-9
Sequence 9, Application US/08656034
Patent No. 6015691
Patent No. 6015691 5786172
GENERAL INFORMATION:
APPLICANT: Walker, David H.
APPLICANT: Yu, Xue-Jie
TITLE OF INVENTION: IMMUNODOMINANT 120 KDA SURFACE-EXPOSED
ADHESION PROTEIN GENES OF EHRlichia

TITLE OF INVENTION: CHAFFEENSIS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/656,034
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Cordeir, Timothy S.
REGISTRATION NUMBER: 38,414
REFERENCE/DOCKET NUMBER: UTSG:191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1716 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 57..1700
US-08-656-034-9

Query Match 1.8%; Score 36.4; DB 3; Length 1716;
Best Local Similarity 53.5%; Pred. No. 0.12;
Matches 76; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

633 TAGCAACCACTTGTGAATATGACATATGAGGCTATGCGGCTTCTTTTGGACAGA 692
DB 149 TAATAGATATTTGTCACACACAGATCTGGGTTCTGAATTTATCAGACAGCAAGA 208
693 AAAGAACCAAGCTATGCTGCTGACAGAGAGCCGTTGTAGAGCGCTTCTTCATCAA 752
DB 209 AGATGACACAGCTCTCAACCTTCATTTAGAGCATTTGTGAGAAAGTATTTCTAA 268
753 TCTGAGAGACGTGTACCCAC 774
DB 269 AGTTGACACAGAAAAAATAAC 290

RESULT 7

US-08-656-034-1
Sequence 1, Application US/08656034
Patent No. 6015691
Patent No. 6015691 5786172
GENERAL INFORMATION:
APPLICANT: Walker, David H.
APPLICANT: Yu, Xue-Jie
TITLE OF INVENTION: IMMUNODOMINANT 120 KDA SURFACE-EXPOSED
ADHESION PROTEIN GENES OF EHRlichia
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:

Db 1032 GAGTGTTCGATCAACCTTAACATATGACCAACATTTAAGAGATTATGAAAAATTAA 1091
QY 659 ATATGG 664
Db 1092 AGAAGC 1097

RESULT 10

US-08-257-073-8
Sequence 8, Application US/08257073
Patent No. 5766597

GENERAL INFORMATION:

APPLICANT: Proletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Jine, John A.
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtiss, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
CLASSIFICATION: 424
FILING DATE: 09-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1884 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-257-073-8

Query Match 1.6%; Score 33.2; DB 1; Length 1884;
Best Local Similarity 54.0%; Pred. No. 1.6; Mismatches 58; Indels 0; Gaps 0;

QY 539 ACAAGGAATGCTGTGTTGAAGTGTGATTTGAATGAGAAAGACTTGGCTTATCA 598
Db 972 ACATGTAATGATTTTCACGACATTTGATCTTTTGAATGATATAATAGTTTGAATT 1031
QY 599 GCTGTTGTCATCCACCCGCTGAGCATGAACAATAGCAACCACTTGTGAATAGTACA 658
Db 1032 GAGTGTTCGATCAACCTTAACATATGACCAACATTTAAGAGATTATGAAAAATTAA 1091
QY 659 ATATGG 664

Db 1092 AGAAGC 1097

RESULT 11

US-09-134-001C-2607
Sequence 2607, Application US/09134001C
Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 2607
LENGTH: 1554
TYPE: DNA
ORGANISM: Staphylococcus epidermidis

US-09-134-001C-2607

Query Match 1.6%; Score 33; DB 4; Length 1554;
Best Local Similarity 46.1%; Pred. No. 1.7;
Matches 111; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 1518 TGCAATACATCCACATCTTCAGAGACCTTGCTAAGCAATGCTGACGACAGTAGCCAA 1577
Db 726 TGAATTCGTGCAACCGTTTATAGTTGAAGCCAAACATGCTGGAACGATGAGACGC 785
QY 1578 CATGACAGTGTGTCGCCAGCCCTCATCTTCAAAAATCTAATAGCTTACTATTCCTAA 1637
Db 786 AGTATCTTTTATATTTCCGTCAGATAGACACTGACATCTTGAAGATCTTACTAA 845
QY 1638 TATTCGGATAGAGCACCACATTTGGCAAAACATTTGGGATGCTTGCAGCCAGCAAAAC 1697
Db 846 TAAACGCTTGTGATGATTTAAAGTTGAACAAGTGAACAACCTATCTACGCTACATTCAC 905
QY 1698 TGAATACAGAGTGGAGAGAGTGTGATATATTTAGGCGCTAACCCAAAGATTC 1757
Db 906 GAAATATATGACAAATGTAGATGCTGAATTTGTATTTGAAAAAGTTGACTTAATATATAC 965
QY 1758 A 1758
Db 966 A 966

RESULT 12

US-09-011-745-1

Sequence 1, Application US/09011745
Patent No. 6165715

GENERAL INFORMATION:

APPLICANT: Collins, Mary KL
APPLICANT: Weiss, Robin A
APPLICANT: Takeuchi, Yasuhiro
APPLICANT: Cosset, Francois-Loic
TITLE OF INVENTION: Expression systems
FILE REFERENCE: 09/011,745
CURRENT APPLICATION NUMBER: US/09/011,745
CURRENT FILING DATE: 1998-06-22
EARLIER APPLICATION NUMBER: PCT/GB96/02061
EARLIER FILING DATE: 1996-08-23
EARLIER APPLICATION NUMBER: GB9517263.1
EARLIER FILING DATE: 1995-08-23
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 2518
TYPE: DNA
ORGANISM: RD114

FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)
OTHER INFORMATION: n is any nucleotide
US-09-011-745-1

Query Match 1.6%; Score 32.6; DB 4; Length 2518;
Best Local Similarity 51.0%; Pred. No. 3.2;
Matches 77; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 911 GGATCATCTATCAGACGCCAGAGAGCTGTATGCTTGCCTCCAGAGAGTGACCGGCC 970
DB 1947 GGACCTTACTAAGCGGCAACAGAGAGGAAATTTGTTAGCTTACAAAGAAAATGCTGTTT 2006
QY 971 CAGTGTTCGAGCTCACCAGTGGGTGAACATGACAGATGTGAGCGTCCAGCTCAATGCCA 1030
DB 2007 TTATGCTTACAGTCAAGTCAAGATTTGTGAGAAACAAATATAGAACCTTACAGAGATTTACA 2066
QY 1031 CACACACGTGAAGAGCTGTAAACCTGCCT 1061
DB 2067 AAAACGCGAGGAAAGCCTGGCAACCAACCT 2097

RESULT 13
US-09-011-745-8
Sequence 8, Application US/09011745
Patent No. 6165715
GENERAL INFORMATION:
APPLICANT: Collins, Mary KL
APPLICANT: Weiss, Rodin A
APPLICANT: Takeuchi, Yasuhiro
APPLICANT: Cosset, Francois-Joic
TITLE OF INVENTION: Expression systems
FILE REFERENCE: 09/011,745
CURRENT APPLICATION NUMBER: US/09/011,745
EARLIER FILING DATE: 1998-06-22
EARLIER APPLICATION NUMBER: PCT/GB96/02061
EARLIER FILING DATE: 1996-08-23
EARLIER APPLICATION NUMBER: GB9517263.1
EARLIER FILING DATE: 1995-08-23
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 8
LENGTH: 5865
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Portion of
OTHER INFORMATION: construct
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3611)
OTHER INFORMATION: n is any nucleotide
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3612)
OTHER INFORMATION: n is any nucleotide
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3613)
OTHER INFORMATION: n is any nucleotide
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3614)
OTHER INFORMATION: n is any nucleotide
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3799)
OTHER INFORMATION: n is any nucleotide
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3800)
OTHER INFORMATION: n is any nucleotide

FEATURE:
NAME/KEY: misc_feature
LOCATION: (3801)
OTHER INFORMATION: n is any nucleotide
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3802)
OTHER INFORMATION: n is any nucleotide
US-09-011-745-8

Query Match 1.6%; Score 32.6; DB 4; Length 5865;
Best Local Similarity 51.0%; Pred. No. 5.9;
Matches 77; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 911 GGATCATCTATCAGACGCCAGAGAGCTGTATGCTTGCCTCCAGAGAGTGACCGGCC 970
DB 2805 GGACCTTACTAAGCGGCAACAGAGAGGAAATTTGTTAGCTTACAAAGAAAATGCTGTTT 2864
QY 971 CAGTGTTCGAGCTCACCAGTGGGTGAACATGACAGATGTGAGCGTCCAGCTCAATGCCA 1030
DB 2865 TTATGCTTACAGTCAAGTCAAGATTTGTGAGAAACAAATATAGAACCTTACAGAGATTTACA 2924
QY 1031 CACACACGTGAAGAGCTGTAAACCTGCCT 1061
DB 2925 AAAACGCGAGGAAAGCCTGGCAACCAACCT 2955

RESULT 14
US-08-030-410-2
Sequence 2, Application US/08030410
Patent No. 6221359
GENERAL INFORMATION:
APPLICANT: Komiyama, Atsushi
APPLICANT: Nakahata, Tatsutoshi
APPLICANT: Kubo, Tetsuo
APPLICANT: Tanaka, Ryuhel
APPLICANT: Kawano, Genji
APPLICANT: Sudo, Tetsuo
APPLICANT: Sano, Emiko
APPLICANT: Kojima, Katsuki
TITLE OF INVENTION: HEMOPOIETIC STEM CELL MULTIPLIER
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/030,410
FILING DATE: 19930521
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8898
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2172 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 1..2169
US-08-030-410-2

Query Match 1.6%; Score 32.4; DB 4; Length 2172;
Best Local Similarity 60.0%; Pred. No. 3.4;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 12 CTACATTGGCGCTTGGAGAGCGGATTCACAGACAGTGTGAGATATCATTTGATGGG 71
|||||
DB 1839 CTACACTGGATTGATCACTATGATGGCCCATTTAGAGTGGCAGCATCTCTATATATGSG 1898
|||||

OY 72 CTATGGCAAAAATGGCCAGAAATGCAGGGG 101
|||||
DB 1899 AAATGAGAAATGCAGCCAGCATCATCGAGG 1928
|||||

RESULT 15
US-09-465-558-61
Sequence 61, Application US/09465558
Patent No. 6436657

GENERAL INFORMATION:

APPLICANT: Morokino, Lajo O.

FILE OF INVENTION: TETRAHYDROFOLATE METABOLIC ENZYMES

CURRENT APPLICATION NUMBER: US/09/465,558

CURRENT FILING DATE: 1999-12-17

EARLIER APPLICATION NUMBER: 60/112,734

NUMBER OF SEQ ID NOS: 70

SOFTWARE: Microsoft Office 97

SEQ ID NO 61

LENGTH: 1021

TYPE: DNA

ORGANISM: Trillium aestivum

US-09-465-558-61

Query Match 1.6%; Score 31.8; DB 4; Length 1021;
Best Local Similarity 47.3%; Pred. No. 3.1;

Matches 96; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

OY 402 TATAGCTCACACAATCTTAACGAGCAAAATCTTTATCAACAAGAAATGTTGCTAA 461
|||||

DB 136 TTTAGATCATGATGTGAGTGAATGTGGCTCTGTTGAGCTGGATGTTAATCT 195
|||||

OY 462 TGTGAGATCAACGAGCGCCCTCTTACCTTCTGAATCCACAGTCAGAGAGCAAG 521
|||||

DB 196 TGCTAAGCATCTCAACACAAAGGTTATGAGTGAATGTTAGTACATCAACAA 255
|||||

OY 522 GTATTCTTCAACAGACAGAAAGTGTGTTGAACCTGTGATTTGATGAGAGA 581
|||||

DB 256 ATTTCACAGACAGACAGAGCGTGAATGGAAGTTGTGAAGGCGGCTATGCGAGCTGG 315
|||||

OY 582 AGACTTGGGCTTTATCAGCTGT 604
|||||

DB 316 GGCCTTTCATGCTGTGCTCTGCT 338
|||||

Search completed: July 5, 2003, 09:10:41
Job time : 109 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 08:28:18 ; Search time 2979 Seconds

(without alignments)
11139.494 Million cpi1 updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttcaagtgtactactatg.....ttgaagtgtactacttag 2049

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estnu:*
4: em_estnu:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mam:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	683.2	33.3	700	BB356080	BB356080 BB356080
2	670.2	32.7	896	BB237061	BB237061 602027402
3	431	21.0	722	BB553338	BB553338 603193475
4	410.8	20.0	672	AA913512	AA913512 0137h09.s
5	369	18.0	534	BB702763	BB702763 602684640
6	344.4	16.8	813	BB387191	BB387191 601277054

7	322.2	15.7	790	BB459539	BB459539 603200546
8	317.6	15.5	636	BB773189	BB773189 602721925
9	314	15.3	436	BB193951	BB193951 UT-R-CM1-
10	302.2	14.7	484	BB289059	BB289059 602383933
11	253.8	12.4	742	BB153673	BB153673 603190594
12	250.8	12.2	312	BB215231	BB215231 BB215231
13	241.6	11.8	281	BB356081	BB356081 BB356081
14	231.4	11.3	806	BB971187	BB971187 601651171
15	229.6	11.2	789	BB531642	BB531642 601230651
16	228.4	11.1	246	BB160556	BB160556 AV160556
17	202.8	9.9	277	BB242940	BB242940 BB242940
18	200.8	9.8	698	BB327376	BB327376 BB327376
19	200.4	9.8	936	BB560354	BB560354 603253902
20	193.2	9.4	269	BB284970	BB284970 AV284970
21	174	8.5	674	BB201876	BB201876 RPT-23-3
22	164.6	8.0	716	BB1309181	BB1309181 EST530591
23	157	7.7	451	BB220146	BB220146 VY52H04.T
24	155.8	7.6	745	BB602483	BB602483 HVSMEH09
25	151.4	7.4	299	BB556696	BB556696 f555905.Y
26	149.6	7.3	648	BB859350	BB859350 f555905.Y
27	143	7.0	674	BB1447071	BB1447071 NF032C05F
28	143	7.0	733	BB646686	BB646686 EST508305
29	142	6.9	601	BB705497	BB705497 SK49910.Y
30	136.6	6.7	384	BB373225	BB373225 EST85231
31	136.2	6.6	521	BB273133	BB273133 sa017d06.
32	135.6	6.6	641	BB1532035	BB1532035 SD03384.5
33	135.2	6.6	436	BB817265	BB817265 BB817265
34	133	6.5	569	BB462326	BB462326 EST324591
35	131.8	6.4	693	BB1365355	BB1365355 RES0638.5
36	131	6.4	701	BB640328	BB640328 SD11318.5
37	130.4	6.4	619	BB1543512	BB1543512 SD10303.5
38	129.2	6.3	634	BB1098334	BB1098334 IPL 31.B0
39	128.6	6.3	759	BB0850241	BB0850241 QGB12A24.
40	127.4	6.2	834	BB049419	BB049419 RPT-24-3
41	124.2	6.1	688	BB0001648	BB0001648 QGG8G21.
42	124	6.1	562	BB1534650	BB1534650 SD07582.5
43	118.8	5.8	550	BB323458	BB323458 BB323458
44	118.8	5.8	720	BB012999	BB012999 OG3H13.Y
45	118.2	5.8	594	BB1639271	BB1639271 SD21820.5

ALIGNMENTS

RESULT 1
LOCUS BB356080 700 bp mRNA linear EST 24-OCT-2001
DEFINITION BB356080 RIKEN full-length enriched, adult male corpus striatum Mus musculus cDNA clone C030011021 3', mRNA sequence.
ACCESSION BB356080
VERSION BB356080.2 GI:16405735
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 700)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M., and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jul 12, 2000 this sequence version replaced gi:9067908.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

source

1. .896
/organism="Mus. musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="162513"
/clone_lib="NCI_CGAP_L19"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; site: 1; Not:
site: 2; Sali: Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by life
technologies. Note: this is a NCI_CGAP library."

BASE COUNT
ORIGIN

236 a 223 c 229 g 208 t

Query-Match
Best Local Similarity 93.1%; Pred. No. 2.3e-192;
Matches 779; Conservative 0; Mismatches 48; Indels 10; Gaps 7;
Score 670.2; DB 12; Length 896;
1 TGAACCTGGTGAATGATGAGAGACTTGGCTTATGAGCTGTTCATCCATCAC 616
557 TGAACCTGGTGAATGATGAGAGACTTGGCTTATGAGCTGTTCATCCATCAC 616
1 TGAACCTGGTGAATGATGAGAGACTTGGCTTATGAGCTGTTCATCCATCAC 60
617 CCGTGAAGATGACAAATGACAACTTTGATGATGACAAATGACAACTTTGATG 676
61 CCGTGAAGATGACAAATGACAACTTTGATGATGACAAATGACAACTTTGATG 119
677 ACCCTTTGAGCAAGAAAGCAAGGCTATGCTGAGAGAGAGAGAGAGAGAGAG 736
120 ACCCTTTGAGCAAGAAAGCAAGGCTATGCTGAGAGAGAGAGAGAGAGAGAG 179
737 GCTTTCCTTCATCAATCTCGAGAGAGCTGTACCCCAATCTTGGCCGATGTC 796
180 GCTTTCCTTCATCAATCTCGAGAGAGCTGTACCCCAATCTTGGCCGATGTC 239
797 ACACAGGGAGAGCTTGTGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 856
240 ACACAGGGAGAGCTTGTGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 299
857 GCATGGCCAG 916
300 GCATGGCCAG 359
917 TCTATCAG 976
360 TCTATCAG 419
977 TTGACAGCTCAG 1036
420 TTGACAGCTCAG 479
1037 CAGTGAAG 1096
480 CAGTGAAG 539
1097 TTTCGGGCTCAATATTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1156
540 TTTCGGGCTCAATATTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
1157 GGGACAGAGCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1215
600 GGGACAGAGCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659
1216 ATCCGCTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1272
660 ATCCGCTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
1273 GATTCAG 1328
720 GATTCAG 779
1329 AACACAGATGTC-GGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1384
780 ACCACCATGTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 836

RESULT 3

BI553338 722 bp mRNA linear EST 05-SEP-2001

LOCUS

603193475F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5264635 5'

DEFINITION

mRNA sequence.

ACCESSION

BI553338

VERSION

BI553338.1 GI:15440650

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 722)

AUTHORS

NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

FEATURES

Tissue Procurement: Miklos Palokovits, M.D., Ph.D.

SOURCE

CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

LOCATION/Qualifiers

1. 722

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="5264635"

/clone_lib="NIH_MGC_95"

/lab_host="DH10B"

/note="Organ: brain; Vector: pBluescript (modified

pBluescript KS+); Site: 1; BamHI; Site: 2; Sali-XhoI (gtcgag

); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTT-3'

size-selected for

average insert size 2.5 kb and

normalized to

for full-length clones and constructed using the

Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIH/NHGRI, National

Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 214 a 154 c 168 g 186 t

ORIGIN

Query Match 21.0%; Score 431; DB 13; Length 722;

Best Local Similarity 84.5%; Pred. No. 1.7e-119;

Matches 496; Conservative 0; Mismatches 90; Indels 1; Gaps 1;

1463 ACCAGGCTCAGCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1522
47 AACAGGCTCAGCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 106
1523 ACATCAACCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1582
107 ACATCAACCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 166
1583 GCAAGTGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1642
167 GCAAGTGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 226
1643 CGGATAG 1702
227 TGGATAG 1762
1703 ACAGAGTGGAG 1762

Accession	Version	KeyWords	Source	Organism	Reference	Authors	Title	Journal	Comment
BI459539	1	GI:15250195	human.	Homo sapiens					
BI459539.1	1	GI:15250195	human.	Homo sapiens					
EST.									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									

QY	1883	TGGGCTAGCAATGCACCAAC--TATACTGGCAATTCACATATCTGCTTACCCTGGAAAT	1940
DB	717	TGGGCTTAGTAATATGCACACAGCTGGGATGGCCCTATTTCACACACTGGCCACGCTTGAAAT	776
QY	1941	CTACAGAAATTAAGAT	1954
DB	777	CTACAGAAATCAGAT	790
RESULT 8			
Bg773189	Bg773189	636 bp	linear
LOCUS	602721925P1	NIH_MGC_97	Homo sapiens
DEFINITION	NIH_MGC_97	Homo sapiens	cdna clone IMAGE:4838671 5', mRNA sequence.
ACCESSION	Bg773189		
VERSION	Bg773189.1	GI:14083842	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 636)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: c9apbs-remail.nih.gov		
	Tissue Procurement: Miklos Rakovits, M.D., Ph.D.		
	CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki Toshlyuki and Piero Carninci (RIKEN)		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:		
	http://linalg.lnl.gov		
	Plate: LNLMI0773	row: a	column: 08
	High quality sequence stop: 636.		
FEATURES			
source	Location/Qualifiers		
	1..636		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_image="4838671"		
	/clone_idb="NIH_MGC_97"		
	/lab_host="DH10B"		
	/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site.1: BamH; Site.2: SalI-XhoI (gtcggag)"; Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTIVN-3', size-selected for average insert size 2.2 kb and normalized to R0T 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH-MGC Library."		
BASE COUNT	183 a	131 c	157 g
ORIGIN			165 t
Query Match	15.5%;	Score 317.6;	DB 12; Length 636;
Best Local Similarity	85.7%;	Pred. No. 5.8e-85;	
Matches	353;	Conservative 0;	Mismatches 59; Indels 0; Gaps 0;
QY	1638	TATTGCGGATAGACACCAATTTGGCAACATTTTGGGGATGCTTTGCAGCCAGCAAAACC	1697
DB	156	TATTGTGGATAGACACCAAAAGGACAGAACTTTGGGGATGTCTCTGCAGCAGCAAAACC	215
QY	1698	TGAATACAGAGTGGGAGAAAGTGTTGAAGTTATATTTGTAGCGCTAACCAAAAGATTC	1757
DB	216	TGAATACAGAGTGGGAGAAAGTTGCTGAAGTTATATTTGTAGTCTAACCCGAAGATTC	275
QY	1758	AGCAGAGAACCAAGCCATCAACCTTCCTACGTGGAGAAATACGAGACCTGTAGAC	1817
DB	276	AGTCAAAACCAAGCCATCAAGCTTCCTCTACTGTGGAGAAATATGAGGCTACTTCAAC	335

```

QY      1818 TGAATGAGCAATATGATACGATGCTCTGGAGAGAGAGGTTTATGGACAAAGG 1877
        |||||||  |||  |||  |||||||  |||  |||  |||||||  |||  |||  |||
Db      336 ATGCTGGCAGATAGTGTATGATGATCCCTCTGGAGACTCGTTTATGGACAAAGG 395
QY      1878 AATAGTGGCTGAGCAATCAACAATATAGTGGCATATTCAGATACCTACCTGG 1937
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      396 ACTCCGCGGTGAGTAATGCAACAGTGAATGGCATATTCAGACACATCCAGCCTGG 455
QY      1938 AATCTACAGATATGATTTGGACATCGAAGCAACTCTGTAACCCGCTGT 1997
        |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||
Db      456 AATCTACAGATATGATTTGGACATCGAAGCAACTCTGTAAGCCTGTGT 515
QY      1998 CATACAGATTTGAAGCAATTTCTCTCTTTGAAGTGTGACTACTTGA 2049
        |||||||  |||||||  |||  |||  |||  |||  |||  |||  |||  |||
Db      516 CATACATTTGTAAGGCACTTCCCGGCTTTGTAAGTGTGACTACTTGA 567

RESULT 9
LOCUS   B0193951 436 bp mRNA linear EST 30-APR-2002
DEFINITION UI-R-CNI-cliz-o-04-0-UI s1 UI-R-CNI Rattus norvegicus cDNA clone
ACCESSION B0193951
VERSION   B0193951
KEYWORDS  B0193951.1 GI:20369502
SOURCE    EST.
ORGANISM  Norway rat.
           Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE 1 (bases 1 to 436)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
JOURNAL   discovery
MEDLINE   Genome Res. 6 (9), 791-806 (1996)
COMMENT   97044477
           Contact: Soares, MB
           Program for Rat Gene Discovery and Mapping
           University of Iowa
           451 Eckstein Medical Research Building Iowa City, IA 52242, USA
           Tel: 319 335 8250
           Fax: 319 335 9565
           Email: msoares@blue.weeg.uiowa.edu
           The sequence contained an oligo-dT track that was present in the
           oligonucleotide that was used to prime the synthesis of first
           strand cDNA and therefore this may represent a bonafide poly A
           tail. The sequence tag present in the cDNA between the NotI site
           and the oligo-dT track served to identify it as a clone from the
           normalized fundus library cDNA library Preparation: M.B. Soares Lab
           Clone distribution: clones will be available through Research
           Genetics (www.resgen.com)
           Seq primer: M13 forward
           POLYA-Tes.

FEATURES
source    Location/Qualifiers
           1..436
           /organism="Rattus norvegicus"
           /strain="Sprague-Dawley"
           /db_xref="taxon:10116"
           /clone="UI-R-CNI-cliz-o-04-0-UI"
           /clone_1ib="UI-R-CNI"
           /dev_stage="adult"
           /note="Vector: pT7p3D-Pac (Pharmacia) with a modified
           polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-CNI
           library is a subtracted library derived from the following
           pool of seven normalized rat libraries: normalized rat
           seminal vesicles, normalized rat penis, normalized rat
           bladder, normalized rat cervix, normalized rat brown
           adipose, normalized rat fundus, and normalized rat
           salivary gland. It was constructed according to the
           procedure described by Bonaldo, Lennon & Soares (Genome
           Research Genome 6: 791-806, 1996). For construction of
           the CNI library, plasmid DNA from the pool of normalized

```

libraries was electroporated into competent bacteria for the production of single-stranded circular DNA. This was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a pool of about 34,000 clones from the Rat ungine Set corresponding to plates R-5-AA-NN excluding plates R-5-MM and MN. This pool represented 40% of the final driver population. b) a pool of about 29,000 clones from subtracted libraries C40 and C41 corresponding to plates R-C40-AMV through R-C40-AXS, R-C40-AZX through R-C40-BAL, R-C40-BFE through R-C40-BHY, R-C40-BJS, R-C40-BKE, R-C40-BKG-H, R-C40-BKJ-K, R-C40-BKR through R-C40-BKS, R-C40-BKU-V, R-C40-BLY through R-C40-BMA, R-C40-BMC through R-C40-BME, R-C40-BNS, R-C40-BOB through R-C40-BOI, R-C40-BPA through R-C40-BPG, R-C41-BBA through R-C41-BDA, R-C41-BHZ through R-C41-BJF, R-C41-BJR, R-C41-BJT through R-C41-BKB, R-C41-BKD, R-C41-BKF, R-C41-BKI, R-C41-BKT, R-C41-BLF, R-C41-BLH through R-C41-BLN, R-C41-BLS, R-C41-BLU-V, R-C41-BNR, and R-C41-BLE. The resulting pool represented 20% of the final driver population. c) a pool of about 15,000 clones from non-normalized libraries CS0s, CT0s, CU0s, CW0s, CX0s and normalized libraries CS0, CT0, CU0, CW0, and CX0 corresponding to plates R-CS0s-CBD through R-CS0s-CBO, R-CT0s-CAM through R-CT0s-CAX, R-CU0s-CBP through R-CU0s-CCA, R-CW0s-CCB through R-CW0s-CCM, R-CX0s-CCN through R-CX0s-CCX, R-CS0-BSD, R-CS0-BPD through R-CS0-BTV, R-CS0-BVM, R-CT0-BTW through R-CT0-BOP, R-CT0-BVN, R-CU0-BUQ through R-CU0-BVL, R-CW0-BVV through R-CW0-BMP, R-CX0-BXN through R-CX0-BXO, R-CX0-BMQ through R-CX0-BKM. The resulting pool represented 5% of the final driver population. d) a pool of about 5,000 clones (1,000 from non-normalized eye library CY0 and 4,000 from normalized eye library CV1) corresponding to plates R-CV0-BRH through R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through R-CV1-BTC, and R-CV1-BVO through R-CV1-BVO. This pool represented about 5% of the final driver population. e) A pool of about 10,000 clones from subtracted library BS2, BV0 and BVOP (7-9.5 kb cDNA library fraction from rat whole embryo), and BX0 (0.5-7kb cDNA library fraction from rat whole embryo) corresponding to plates R-BS2-BDB through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BVOP-AOI through R-BVOP-AOX, and R-BX0-AY through R-BX0-ASH. The resulting pool represented 5% of the final driver population. f) a pool of about 7,000 clones from the seven non-normalized libraries that make up the tracer including CY0, CZ0, DA0, DB0, DD0, and DE0 corresponding to plates R-CY0-BXP through R-CY0-BXZ, R-CZ0-BYA through R-CZ0-BYI, R-CZ0-BZB-C, R-DA0-BYJ through R-DA0-BYP, R-DA0-BZD through R-DA0-BZH, R-DB0-BYQ through R-DB0-BZA, R-DC0-BZI through R-DC0-BZO, R-DD0-CAY through R-DD0-CBA, R-DD0-BZR through R-DD0-CAH, R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The resulting pool represented about 10% of the final driver population. g) a pool of about 2,000 clones from the pool of normalized libraries, CN0, that makes up the tracer. The corresponding plates are R-CN0-BXW through R-CN0-BLD, R-CN0-BLE, R-CN0-BLP through R-CN0-BLR, R-CN0-BLF, R-CN0-BLM-X, R-CN0-BMB, and R-CN0-BME through R-CN0-BML. This pool represented 5% of the final driver population. h) a pool of the 28 most abundant clones in the CN0 pool corresponding to the following addresses: bkw-a-09-0-UI, bkw-b-09-0-UI, bkw-b-11-0-UI, bkw-b-10-0-UI, bkw-d-01-0-UI, bkw-d-06-0-UI, bkw-g-08-0-UI, bkw-h-12-0-UI, bky-a-05-0-UI, bky-a-06-0-UI, bky-a-11-0-UI, bky-c-06-0-UI, bky-c-09-0-UI, bky-d-10-0-UI, bla-a-01-0-UI, bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI, blb-a-12-0-UI, blb-f-02-0-UI, blc-a-11-0-UI, blc-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, blg-h-04-0-UI, blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5% of the final driver population. i) One abundant CN0 clone (corresponding to the address bky-a-11-0-UI) was digested with Not I and Eco RI and the resulting insert was gel

purified. This purified insert was added directly to the driver so that it represented 5% of the final driver population.
TAG_Lib-UI-R-CNI
TAG_TISSUE-fundus
TAG_SEQ-TTCGG*

BASE COUNT 107 a 96 c 90 g 143 t
ORIGIN

Query Match 15.3%; Score 314; DB 14; Length 436;
Best Local Similarity 94.2%; Pred. No. 5.4e-84;
Matches 326; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1704 CAGAGTGGAGAGTGGTGAAGTTATATTTGTAGGCGCTAACCCAAAGAAATTCAGCAGA 1763
DB 427 CAGAGTGGAGAGTGGTGAAGTTATATTTGTAGGCGCTAACCCAAAGAAATTCAGCAGA 368
QY 1764 GAACCAAGCCCATCAAACTTCCTCAGCTGTGAGAAATACAGAGACTGTGAGCTGACTG 1823
DB 367 AAACCAAGCCCATCAAACTTCCTCAGCTGTGAGAAATACAGAGACTGTGAGCTGACTG 308
QY 1824 GCAGATATATGTATTAACGATGCTCTGAGAGAGAGAGTATTTATTTGGACAAAGAAATTA 1883
DB 307 GCAGATATATGTATTAACGATGCTCTGAGAGAGAGAGTATTTATTTGGACAAAGAGTACT 248
QY 1884 GGGTCTGAGCAATGCACAAATATTAATGATATTCAGATATTCAGATGCTGAGTACTA 1943
DB 247 GGGTCTGAGCAATGCACAAATATTAATGATATTCAGATATTCAGATGCTGAGTACTA 188
QY 1944 CAGAAATAGATATTTTGGACAAATTCGAGAGAGAGAACTTCTGAACCCGCTGATACT 2003
DB 187 CAGAAATAGATATTTTGGACAAATTCGAGAGAGAGAACTTCTGAACCCGCTGATACT 128
QY 2004 AGCAATTTGAAGAAATTTCTCTCTTTTGAAGTTGCTCACTACTAG 2049
DB 127 AGCAATTTGAAGAAATTTCTCTCTTTTGAAGTTGCTCACTACTAG 82

RESULT 10 484 bp mRNA linear EST 21-FEB-2001
LOCUS BG289059 603383933P1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4512680 5',
DEFINITION mRNA sequence.
ACCESSION BG289059
VERSION BG289059.1 GI:13044522
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 484)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM10398 row: b column: 09
High quality sequence stop: 484.

FEATURES
SOURCE location/Qualifiers
1..484

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4512680"
/clone_lib="NIH_MGC_93"
/tissue_type="translational cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"

/note="Organ: bladder; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; Oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."

BASE COUNT 158 a 93 c 107 g 126 t
ORIGIN

Query Match 14.7%; Score 302.2; DB 12; Length 484;
Best Local Similarity 85.3%; Pred. No. 2.3e-80;
Matches 337; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1635 CAATTGGCAACATTTTGGGATGCTTCGACCCAGCAAAACCTGATATACAGAGTGGAG 1714
DB 1 CAAAAGCGAGAACTTCGCGGATGCTTCGACCCAGCAAAACCTGATATACAGAGTGGAG 60
QY 1715 AAGTGTGAAGTTATTTGAGCGGCTAACCCAAAGAAATTCAGAGAGAGAGAGCC 1774
DB 61 AAGTGTGAAGTTATTTGAGCGGCTAACCCAAAGAAATTCAGAGAGAGAGAGAGAGCC 120
QY 1775 ATCAAACTTCCTCAGCTGTGAGAAATACAGAGACTGTGAGCTGAGAGATATATGT 1834
DB 121 ATCAAACTTCCTCAGCTGTGAGAAATACAGAGACTGTGAGCTGAGAGATATATGT 180
QY 1835 ATACAGATGCTCTCGGAGAGAGAGTATTTATTTGACAAAGAAATATGAGTGTGAGCA 1894
DB 181 ATACAGATGCTCTCGGAGAGAGAGTATTTATTTGACAAAGAGAGTGTGAGTGTGAGCA 240
QY 1895 ATGCAACATATATGAGATATTCAGATATTCAGATATTCAGATATTCAGATATTA 1954
DB 241 ATGCAACATATATGAGATATTCAGATATTCAGATATTCAGATATTCAGATATTA 300
QY 1955 ATTTTGGACAAATTCGAGAGAGAGAACTTCGAAACCCGCTGATACTAGATTGAG 2014
DB 301 ATTTTGGACAAATTCGAGAGAGAGAACTTCGAAACCCGCTGATACTAGATTGAG 360
QY 2015 GAATTTCTTCCTTTTGAAGTTGCTCACTACTAG 2049
DB 361 GAATTTCTTCCTTTTGAAGTTGCTCACTACTAG 395

RESULT 11 742 bp mRNA linear EST 05-SEP-2001
LOCUS B1553673 603190594P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262094 5',
DEFINITION mRNA sequence.
ACCESSION B1553673
VERSION B1553673.1 GI:15440985
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 742)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
Toshiyuki and Piero Carninci (RIEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM11660 row: k column: 23
High quality sequence stop: 734.

FEATURES
SOURCE location/Qualifiers
1..742

/organism="Homo sapiens"
/db_xref="taxon:9606"

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 07:05:53 ; Search time 468 Seconds
(without alignment)
9859.711 Million cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttcaagtgcactactacatggtg.....ttgaagtgtcactactag 2049

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2049	100.0	2049	21	AAA97634
2	2049	100.0	2271	21	AAA97633
3	2049	100.0	3108	21	AAA97632
4	2047.4	99.9	4835	21	AAA97635
5	1830	89.3	2283	22	AA99977
6	1511.4	73.8	2507	22	AA22668
7	524.8	25.6	351	22	AAH98033
8	441.4	21.5	3058	24	AB19947
9	431.8	21.1	1194	20	AA206239

10	396.4	19.3	539	22	AA99979	PCR product isolat
11	334.8	16.3	2553	23	AB102671	Drosophila melanog
12	166.4	8.1	289	21	AA03574	Human secreted pro
13	155.2	7.6	7083	23	AB102670	Drosophila melanog
14	121.2	5.9	1941	21	AA290566	P. aeruginosa matu
15	121.2	5.9	2013	21	AA290573	P. aeruginosa cera
16	80	3.9	4411529	22	AA199682	Mycobacterium tube
17	77.6	3.8	565	22	AA199682	Drosophila melanog
18	74.2	3.6	4403765	22	AA199683	Mycobacterium tube
19	65	3.2	65	24	ABN3673	Mouse spliced tran
20	47.2	2.3	60	24	ABN36755	Human spliced tran
21	40.4	2.0	14041	22	AAH48024	Internal control B
22	38	1.9	38	21	AA97627	Mouse neutral/alka
23	37.4	1.8	462	24	AB165747	Lung cancer relate
24	37.2	1.8	5535	21	AA70184	Plasmodium falcipa
25	36.4	1.8	1920	19	AAV03655	Ethiopia chaffee
26	36.2	1.8	2258	23	AA584963	DNA encoding novel
27	35.4	1.7	2258	23	AB104267	Drosophila melanog
28	35.4	1.7	4324	23	AB104266	Drosophila melanog
29	35.4	1.7	5460	24	ABK51683	Human ABCG5 upstre
30	34.8	1.7	38	24	AA22672	Human ceramidase c
31	34.8	1.7	1716	19	AAV03656	Ethiopia chaffee
32	34.8	1.7	2041	23	AA583709	DNA encoding novel
33	34.8	1.7	2998	23	AB109036	Drosophila melanog
34	34.8	1.7	3293	23	AB109000	Drosophila melanog
35	34.6	1.7	10732	21	AA10594	Gene encoding a su
36	34.4	1.7	4590	22	AAH24065	Yeast AOD604-asso
37	34.2	1.7	654	20	AAV61836	B. burgdorferi ant
38	34.2	1.7	657	20	AAV61836	B. burgdorferi ant
39	34.2	1.7	717	20	AAV61836	B. burgdorferi ant
40	34.2	1.7	783	20	AAV61836	B. burgdorferi ant
41	34.2	1.7	910715	20	AA20248	Borrelia burgdorfe
42	34	1.7	582	21	AA98442	Dog PTEN/MAA1 Exo
43	34	1.7	701	21	AA98490	Dog PTEN/MAA1 Int
44	34	1.7	1317	23	ABV22520	Human prostate exp
45	34	1.7	1317	23	ABV28335	Human prostate exp

ALIGNMENTS

RESULT 1	AAA97634	standard; DNA: 2049 BP.
ID	AAA97634	
AC	AAA97634:	
XX		
XX	02-FEB-2001	(first entry)
XX		
DE	Mouse mature neutral/alkaline ceramidase-encoding DNA, SEQ ID NO:15.	
XX		
KW	Neutral/alkaline ceramidase; mouse; murine; recombinant production;	
KW	cellular ceramidase content control; antibody; membrane synthesis;	
KW	lipid engineering; ceramide metabolism; drug development;	
KW	mature ceramidase; ds.	
XX		
OS	Mus sp.	
XX		
PN	W0200058448-A1.	
XX		
PD	05-OCT-2000.	
XX		
PF	24-MAR-2000; 2000MO-JF01802.	
XX		
XX		
PR	26-MAR-1999; 99JP-0084743.	
XX		
PA	(TAKI) TAKARA SHUZO CO LTD.	
XX		
PI	Ito M.	
XX		
DR	WPI: 2000-619079/59.	
XX	P-PSDB; AAB23292.	

PT Mammalian neutral/alkaline ceramidase applicable as lipid engineering
PT reagent for studying structure and functions of ceramide as well as
PT developing drugs for diseases associated with ceramide metabolism
XX
PS Claim 1; Page 65-66; 76pp; Japanese.

XX The invention relates to a mouse neutral/alkaline ceramidase and to
CC nucleic acids encoding it. The invention also relates to expression
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
CC ceramidase, the recombinant production of the ceramidase, an antibody
CC against the ceramidase, and a method for controlling the ceramide
CC content in cells and/or tissues by introducing the gene or its antisense
CC nucleic acid into the cells and/or tissues. The ceramidase can be used
CC as a lipid engineering reagent for studying structure and functions of
CC ceramide as well as developing drugs for diseases associated with
CC ceramide metabolism. The present sequence represents DNA encoding the
CC mature mouse neutral/alkaline ceramidase.

XX
SQ Sequence 2049 BP; 589 A; 470 C; 490 G; 500 T; 0 other:

Query Match 100.0%; Score 2049; DB 21; Length 2049;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGAGTGGCTACTACTGCTGGGAGAGCGGATTGCACAGACAAAGTCTCATATC 60
DB 1 TTGAGTGGCTACTACTGCTGGGAGAGCGGATTGCACAGACAAAGTCTCATATC 60
OY 61 AATTGATGGCTATGCGCAAAATATGGCCAGATGACAGGGGCTCTCACCAGGCTTTC 120
DB 61 AATTGATGGCTATGCGCAAAATATGGCCAGATGACAGGGGCTCTCACCAGGCTTTC 120
OY 121 AGCGGCTCTTTATCTTGGGGGATCCAGATGGGTCAATATGCAATGCAATTTGAGCGTG 180
DB 121 AGCGGCTCTTTATCTTGGGGGATCCAGATGGGTCAATATGCAATGCAATTTGAGCGTG 180
OY 181 GAACATATGATGATTTCCCAACGACGTAGGTTGGAGGTCCTGAGAGAGCTAGAGATAA 240
DB 181 GAACATATGATGATTTCCCAACGACGTAGGTTGGAGGTCCTGAGAGAGCTAGAGATAA 240
OY 241 TATGCTCTCTGTATCGAAGAGACAAATGTTATCCTGATGGTCCATTACACACTCTGGC 300
DB 241 TATGCTCTCTGTATCGAAGAGACAAATGTTATCCTGATGGTCCATTACACACTCTGGC 300
OY 301 CCAGCAGGGTTTTCATATACACTCTATATCTCCGACGAGGATTCAGCAACCGG 360
DB 301 CCAGCAGGGTTTTCATATACACTCTATATCTCCGACGAGGATTCAGCAACCGG 360
OY 361 ACCTTCAGTACATAGTCTCTGGGATCATGAAAGCAATGATATAGCTCACACAAATCTT 420
DB 361 ACCTTCAGTACATAGTCTCTGGGATCATGAAAGCAATGATATAGCTCACACAAATCTT 420
OY 421 AAACGAGGCAAAATCTTTATCAACAAAGAAATGTTCTATATGTCAGATCAACCGAAGC 480
DB 421 AAACGAGGCAAAATCTTTATCAACAAAGAAATGTTCTATATGTCAGATCAACCGAAGC 480
OY 481 CCCCTCTTACCTTCTGGAATCCACAGTCAAGAGACAGATTTCTTCAACACAGAC 540
DB 481 CCCCTCTTACCTTCTGGAATCCACAGTCAAGAGACAGATTTCTTCAACACAGAC 540
OY 541 AAGGAATGCTGCTGCTGAAACTGGTGAATTTGAGAGAGACTTGGGCTCTATAGC 600
DB 541 AAGGAATGCTGCTGCTGAAACTGGTGAATTTGAGAGAGACTTGGGCTCTATAGC 600
OY 601 TGGTTTGCATCCACCCCGTAGCATGAACATAGCAACCTTTGGAATAGGACAAT 660
DB 601 TGGTTTGCATCCACCCCGTAGCATGAACATAGCAACCTTTGGAATAGGACAAT 660
OY 661 ATGGGCTATCGGGCTTACCTTTTGGACAGAAAGAAAGCAAGCTATCTGCTGGACAG 720
DB 661 ATGGGCTATCGGGCTTACCTTTTGGACAGAAAGAAAGCAAGCTATCTGCTGGACAG 720
OY 721 GAGCCGTTTGTAGACAGGCTTGTCTTCAAAATCTCGAGACGTGTCACCAACATCTT 780

DB 721 GAGCCGTTTGTAGACAGGCTTGTCTTCAAAATCTCGAGACGTGTCACCAACATCTT 780
OY 781 GGGCCGATTTGTCAACACAGGAGAGCTTTGTGACAAAGCAAGACACCTGTCCCAAC 840
DB 781 GGGCCGATTTGTCAACACAGGAGAGCTTTGTGACAAAGCAAGACACCTGTCCCAAC 840
OY 841 GGTGGCTTACATGTCATGTCAGGAGGACCTTGACACAGATGTTTGGACACACAC 900
DB 841 GGTGGCTTACATGTCATGTCAGGAGGACCTTGACACAGATGTTTGGACACACAC 900
OY 901 ATTATGAGGATCATCTATACAGAGCCAAAGAGGTATGCTGCTGCCCTCCAGAG 960
DB 901 ATTATGAGGATCATCTATACAGAGCCAAAGAGGTATGCTGCTGCCCTCCAGAG 960
OY 961 GTGACGGCCGAGTCTTGGACGTACACAGTGGGTGAACATGACAGATGTGAGGTCAG 1020
DB 961 GTGACGGCCGAGTCTTGGACGTACACAGTGGGTGAACATGACAGATGTGAGGTCAG 1020
OY 1021 CTCAATGCGCACACACAGTGAAGAGCTGTAAACCTGCTGGGCTACAGTTTCCGCA 1080
DB 1021 CTCAATGCGCACACACAGTGAAGAGCTGTAAACCTGCTGGGCTACAGTTTCCGCA 1080
OY 1081 GGCACAAATGATGAGTGGGCTCAATATTTACAGAGGAACTACGGAAGGGATCCA 1140
DB 1081 GGCACAAATGATGAGTGGGCTCAATATTTACAGAGGAACTACGGAAGGGATCCA 1140
OY 1141 TTTGAGACACTCTTGGGACACAGCTCTTGGGAAACCATCTGAAAGATTTGAGAGTGT 1200
DB 1141 TTTGAGACACTCTTGGGACACAGCTCTTGGGAAACCATCTGAAAGATTTGAGAGTGT 1200
OY 1201 CAGAAACCCCAACCAATCTCTCTACAGTGGAGAGCTGACAGATACCAATCTTGGCAA 1260
DB 1201 CAGAAACCCCAACCAATCTCTCTACAGTGGAGAGCTGACAGATACCAATCTTGGCAA 1260
OY 1261 CCAGATATTTGTTGATGTTGATGATGTTACAGATTTGATGATGTTGATGATGATGAT 1320
DB 1261 CCAGATATTTGTTGATGTTGATGATGTTACAGATTTGATGATGTTGATGATGATGAT 1320
OY 1321 GGGGAATTAACAAACATGTCGGGACGAGAGATTTCTGAGGCAATTAAGAAATTTGCA 1380
DB 1321 GGGGAATTAACAAACATGTCGGGACGAGAGATTTCTGAGGCAATTAAGAAATTTGCA 1380
OY 1381 CTTTATGGGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
DB 1381 CTTTATGGGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
OY 1441 TACATTAACACATATGAAGATACAGAGCTGACAGGCTGACAGAGCATCTACATCTAT 1500
DB 1441 TACATTAACACATATGAAGATACAGAGCTGACAGGCTGACAGAGCATCTACATCTAT 1500
OY 1501 GGACCAACACACCTGTCTGCATACATCAACTCTTCAAGAGCACTTGAAGGCAATGCT 1560
DB 1501 GGACCAACACACCTGTCTGCATACATCAACTCTTCAAGAGCACTTGAAGGCAATGCT 1560
OY 1561 ACGGACACAGTACGACATGAGAGTGGTCCGAGGCTCATCTCTCAAAATCTATATA 1620
DB 1561 ACGGACACAGTACGACATGAGAGTGGTCCGAGGCTCATCTCTCAAAATCTATATA 1620
OY 1621 GCTTCACCTTATTTCTTAATTTGCGATGAGACACAAATTTGGCAAAATTTTGGGATGTC 1680
DB 1621 GCTTCACCTTATTTCTTAATTTGCGATGAGACACAAATTTGGCAAAATTTTGGGATGTC 1680
OY 1681 TTTGAGCCAGCAAAACCTGATATGAGAGTGGGAGAGTGGTGAAGTTATTTGTTAGGC 1740
DB 1681 TTTGAGCCAGCAAAACCTGATATGAGAGTGGGAGAGTGGTGAAGTTATTTGTTAGGC 1740
OY 1741 GCTAACCAAAAGATTTGAGAGAGACAGACCCATTAACCTTCTGACAGTGGAGAAA 1800
DB 1741 GCTAACCAAAAGATTTGAGAGAGACAGACCCATTAACCTTCTGACAGTGGAGAAA 1800
OY 1801 TACGAGACTCTGTAGCTGATGAGATATGATTAACGATGCTCTGGAGACGAGG 1860

Db 1801 TACGAGACTCTGTAGCTGACTGSCAGATATATGTAACGATTCCTCTGGAGACGAGG 1860
Qy 1861 TTTATTGGCACAAGGAATACCTGGTGTGTGAGCAATGCAATATATCTGGCATATTCOA 1920
Db 1861 TTTATTGGCACAAGGAATACCTGGTGTGTGAGCAATGCAATATATCTGGCATATTCOA 1920
Qy 1921 GATPACTGCTACCTGGAATCTACAGATATATTTTGGACAAATCGGAGCAGGAA 1980
Db 1921 GATPACTGCTACCTGGAATCTACAGATATATTTTGGACAAATCGGAGCAGGAA 1980
Qy 1981 CTCTGAAACCCGCTGTCTATCTAGCATTTGAGGAATTTCTCTTTTGAAGTTGTC 2040
Db 1981 CTCTGAAACCCGCTGTCTATCTAGCATTTGAGGAATTTCTCTTTTGAAGTTGTC 2040
Qy 2041 ACTACTTAG 2049
Db 2041 ACTACTTAG 2049
RESULT 2
ID AAA97633 standard; DNA; 2271 BP.
AC AAA97633:
XX
XX
XX 02-FEB-2001 (first entry)
XX
XX Mouse neutral/alkaline ceramidase ORF, SEQ ID NO:1.
XX
XX Neutral/alkaline ceramidase; mouse; murine; recombinant production;
KW cellular ceramide content control; antibody; membrane synthesis;
KW lipid engineering; ceramide metabolism; drug development;
KM open reading frame; ORF; ds.
XX
XX Mus sp.
OS
XX WO200058448-A1.
XX
XX 05-OCT-2000.
XX
XX 24-MAR-2000; 2000WO-JP01802.
XX
XX 26-MAR-1999; 99JP-0084743.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX Ito M;
XX
XX WPI: 2000-619079/59.
XX
XX P-PSDB; AAB23291.
XX
XX Mammalian neutral/alkaline ceramidase applicable as lipid engineering
PT reagent for studying structure and functions of ceramide as well as
PT developing drugs for diseases associated with ceramide metabolism
XX
XX Example 4: Page 57-58; 76pp; Japanese.
XX
XX The invention relates to a mouse neutral/alkaline ceramidase and to
CC nucleic acids encoding it. The invention also relates to expression
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
CC ceramidase, the recombinant production of the ceramidase, an antibody
CC against the ceramidase, and a method for controlling the ceramide
CC content in cells and/or tissues by introducing the gene or its antisense
CC nucleic acid into the cells and/or tissues. The ceramidase can be used
CC as a lipid engineering reagent for studying structure and functions of
CC ceramide as well as developing drugs for diseases associated with
CC ceramide metabolism. The present sequence represents the mouse
CC neutral/alkaline ceramidase open reading frame (ORF).
XX
XX Sequence 2271 BP; 647 A; 541 C; 525 G; 558 T; 0 other:
SO

Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TTCAGTGGCTCTCTACATTGGCGTGGAGACCGGATTCACAGACAAAGTGCATATC 60
Db 223 TTCAGTGGCTCTCTACATTGGCGTGGAGACCGGATTCACAGACAAAGTGCATATC 282
Qy 61 AATTTGATGGGCTATGGCAAAAATATGCGCAGAAATCAGGCGTCTCTACAGGCTTTC 120
Db 283 AATTTGATGGGCTATGGCAAAAATATGCGCAGAAATCAGGCGTCTCTCTACAGGCTTTC 342
Qy 121 AGCCGTCCTTTATCTTGGCGGATTCAGATGGGTCAATATGCATTTTGAAGCGTG 180
Db 343 AGCCGTCCTTTATCTTGGCGGATTCAGATGGGTCAATATGCATTTTGAAGCGTG 402
Qy 181 GAATATGATATGTTTCCACAGCTAGGTTGGAGGTCCTGAGACATGAGATGAA 240
Db 403 GAATATGATATGTTTCCACAGCTAGGTTGGAGGTCCTGAGAGCTGAGATGAA 462
Qy 241 TATGCTCTCTGTATGCAAGAGACATGTTATCCTGATGCCATTTCACACACTCTGCG 300
Db 463 TATGCTCTCTGTATGCAAGAGACATGTTATCCTGATGCCATTTCACACACTCTGCG 522
Qy 301 CCAGCAGGCTTTTCCAAATATATACACTATATCTGCGCAGAGGATTCAGACACCGG 360
Db 523 CCAGCAGGCTTTTCCAAATATATACACTATATCTGCGCAGAGGATTCAGACACCGG 582
Qy 361 ACCTTTCAATCATAGTCTCTGGGATCATGAAGCAATGATATAGCTCACACAAATCTT 420
Db 583 ACCTTTCAATCATAGTCTCTGGGATCATGAAGCAATGATATAGCTCACACAAATCTT 642
Qy 421 AACACGAGCAAAATCTTATCAACAAAGAAATTTGCTAATGTCAGATCAACCGAAGC 480
Db 643 AACACGAGCAAAATCTTATCAACAAAGAAATTTGCTAATGTCAGATCAACCGAAGC 702
Qy 481 CCTCTCTCTTACCTTCTGAATTCACAGTCAAGAGACCAAGATTTCTTAAACACAGAC 540
Db 703 CCTCTCTCTTACCTTCTGAATTCACAGTCAAGAGACCAAGATTTCTTAAACACAGAC 762
Qy 541 AAGGAATGCTGGTCTTGAACCTGTGATTTGAATGAGAGAACTGGGCTTATACGC 600
Db 763 AAGGAATGCTGGTCTTGAACCTGTGATTTGAATGAGAGAACTGGGCTTATACGC 822
Qy 601 TGGTTTGCATCCACCCCGTGAATGCAACCAATGCAACCACTTTGTAATGATGACAT 660
Db 823 TGGTTTGCATCCACCCCGTGAATGCAACCAATGCAACCACTTTGTAATGATGACAT 882
Qy 661 ATGGCTATGCGGCTTACCTTTTGAACAAGAAAGCAAAAGGCTATGCTGAGACAG 720
Db 883 ATGGCTATGCGGCTTACCTTTTGAACAAGAAAGCAAAAGGCTATGCTGAGACAG 942
Qy 721 GGACCGTTTGTAGAGGCTTTGCTTCATCAATCTCGAGACGTTGTACCCACATTTCTT 780
Db 943 GGACCGTTTGTAGAGGCTTTGCTTCATCAATCTCGAGACGTTGTACCCACATTTCTT 1002
Qy 781 GGCCCGATTTGTCAACACAGGAGGAGCTTGTCAACAGCAAGACACACTGTCACAC 840
Db 1003 GGCCCGATTTGTCAACACAGGAGGAGCTTGTCAACAGCAAGACACACTGTCACAC 1062
Qy 841 GGTGGGCTTACATGTCATGTCATGCGGAGCTTGTGAGACATGTTTGAAGACACAC 900
Db 1063 GGTGGGCTTACATGTCATGTCATGCGGAGCTTGTGAGACATGTTTGAAGACACAC 1122
Qy 901 ATTTATGAGCGATATCTATCAAGAGCCCAAGAGCTGTTATGCTCTGCTCCAGAG 960
Db 1123 ATTTATGAGCGATATCTATCAAGAGCCCAAGAGCTGTTATGCTCTGCTCCAGAG 1182
Qy 961 GTGACCGGCGGAGCTTTGCGAGCTACAGAGTGGTGAACATGACAGATGAGAGTCCAG 1020
Db 1183 GTGACCGGCGGAGCTTTGCGAGCTACAGAGTGGTGAACATGAGAGTGGAGTCCAG 1242
Qy 1021 CTCATATGCCACACACAGATGAGAGCTGTAAACCTGCTGGGCTACAGTTTTCGCGCA 1080
Db 1243 CTCATATGCCACACACAGATGAGAGCTGTAAACCTGCTGGGCTACAGTTTTCGCGCA 1302

QY 1081 GGCACAAATGATGAGATTTGGGCTCAATATTACACAGGAATCTACGGAAGGGATCCA 1140
 DB 1303 GGCACAAATGATGAGATTTGGGCTCAATATTACACAGGAATCTACGGAAGGGATCCA 1362
 QY 1141 TTCTGGGACACTTCTGGGACCACTCTGGGAAAAACATCTGAAGAGATTGTAGAGTGT 1200
 DB 1363 TTCTGGGACACTTCTGGGACCACTCTGGGAAAAACATCTGAAGAGATTGTAGAGTGT 1422
 QY 1201 CAGAAACCCAAACCAATCTGCTTCACAGTGGAGACTGACGATATACCAATCTCTGGCAA 1260
 DB 1423 CAGAAACCCAAACCAATCTGCTTCACAGTGGAGACTGACGATATACCAATCTCTGGCAA 1482
 QY 1261 CCAGATATTGTTGATGTTGAGATTGTTACCGTTGGGCTCTGGCCATTCCTATCCCT 1320
 DB 1483 CCAGATATTGTTGATGTTGAGATTGTTACCGTTGGGCTCTGGCCATTCCTATCCCT 1542
 QY 1321 GGGGAATTACAAACCACTGTCGGGACGAAATTTCTGAGAGCAATTTAAAAAGATTTCGA 1380
 DB 1543 GGGGAATTACAAACCACTGTCGGGACGAAATTTCTGAGAGCAATTTAAAAAGATTTCGA 1602
 QY 1381 CTTTATGGGATGAGGATATGACCGTTGTTATCGCAGTCTTAAGCAATGTTTATACACAT 1440
 DB 1603 CTTTATGGGATGAGGATATGACCGTTGTTATCGCAGTCTTAAGCAATGTTTATACACAT 1662
 QY 1441 TACATTATCCACATATGAGAAATACAGGCTCAGCGGTACAGAGCAGCATCTCAATCTAT 1500
 DB 1663 TACATTATCCACATATGAGAAATACAGGCTCAGCGGTACAGAGCAGCATCTCAATCTAT 1722
 QY 1501 GGACCAACACACCTGCTCTGCATACATCCAACTCTTACAGAGCTTCTTAAGCAATTCGT 1560
 DB 1723 GGACCAACACACCTGCTCTGCATACATCCAACTCTTACAGAGCTTCTTAAGCAATTCGT 1782
 QY 1561 ACGGACACAGTACCCAACTAGACAGTGTCTCCGAGCTCTCAATCTCTTCAAAAATCTAATA 1620
 DB 1783 ACGGACACAGTACCCAACTAGACAGTGTCTCCGAGCTCTCAATCTCTTCAAAAATCTAATA 1842
 QY 1621 GCTTACCTTATTCCTTAATATTCGGGATAGAGCAACCAATTTGGCAACATTTTGGGATGTC 1680
 DB 1843 GCTTACCTTATTCCTTAATATTCGGGATAGAGCAACCAATTTGGCAACATTTTGGGATGTC 1902
 QY 1681 TTGCAAGCCAGCAAAACCTGAATATACAGAGTGGGAGAAAGTGTGAAGTTATTTTGAAGC 1740
 DB 1903 TTGCAAGCCAGCAAAACCTGAATATACAGAGTGGGAGAAAGTGTGAAGTTATTTTGAAGC 1962
 QY 1741 GCTAACCCCAAGAAATTCAGCAGAGAAACAGACCCATCAAACTCTCTCACTGTGAGAAA 1800
 DB 1963 GCTAACCCCAAGAAATTCAGCAGAGAAACAGACCCATCAAACTCTCTCACTGTGAGAAA 2022
 QY 1801 TACGAGAGCTCTGTAGCTAGCTAGCGAGATATGTATACGATGCTCTCTGGGAGAGCAGG 1860
 DB 2023 TACGAGAGCTCTGTAGCTAGCTAGCGAGATATGTATACGATGCTCTCTGGGAGAGCAGG 2082
 QY 1861 TTTTATTTGGCACAAGAAAGTACTGGGTGAGCAATGCAAAATATATCTGGCATTTATTCGA 1920
 DB 2083 TTTTATTTGGCACAAGAAAGTACTGGGTGAGCAATGCAAAATATATCTGGCATTTATTCGA 2142
 QY 1921 GATATGCTCTACCTGCTGGAATCTACAGAAATATATTTTGGACACATCTGGAAGCAGAA 1980
 DB 2143 GATATGCTCTACCTGCTGGAATCTACAGAAATATATTTTGGACACATCTGGAAGCAGAA 2202
 QY 1981 CTTTCTGAAAACCGGCTGTCTACTAGCAATTTGAAGCAATTTCTCTCTTTTGAAGTGTTC 2040
 DB 2203 CTTTCTGAAAACCGGCTGTCTACTAGCAATTTGAAGCAATTTCTCTCTTTTGAAGTGTTC 2262
 QY 2041 ACTACTTAG 2049
 DB 2263 ACTACTTAG 2271

RESULT 3
 AAA97632
 ID AAA97632 standard; cDNA; 3108 BP.

XX AC AAA97632;
 XX DT 02-FEB-2001 (first entry)
 DE Mouse neutral/alkaline ceramidase clone, SEQ ID NO:11.
 XX
 KW Neutral/alkaline ceramidase; mouse; murine; recombinant production;
 KW cellular ceramide content control; antibody; membrane synthesis;
 KW lipid engineering; ceramide metabolism; drug development; ss.
 XX
 OS Mus sp.
 PN WO200058448-A1.
 PD 05-OCT-2000.
 PF 24-MAR-2000; 2000MO-JP01802.
 PR 26-MAR-1999; 99JP-0084743.
 PA (TAKI) TAKARA SHUZO CO LTD.
 PI Ito M;
 DR WPI; 2000-619079/59.
 PT Mammalian neutral/alkaline ceramidase applicable as lipid engineering
 PT reagent for studying structure and functions of ceramide as well as
 PT developing drugs for diseases associated with ceramide metabolism -
 PS Example 4; Page 55-56; 76pp; Japanese.
 PS
 CC The invention relates to a mouse neutral/alkaline ceramidase and to
 CC nucleic acids encoding it. The invention also relates to expression
 CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
 CC ceramidase, the recombinant production of the ceramidase, an antibody
 CC against the ceramidase, and a method for controlling the ceramide
 CC content in cells and/or tissues by introducing the gene or its antisense
 CC nucleic acid into the cells and/or tissues. The ceramidase can be used
 CC as a lipid engineering reagent for studying structure and functions of
 CC ceramide as well as developing drugs for diseases associated with
 CC ceramide metabolism. The present sequence represents a clone encoding
 CC mouse neutral/alkaline ceramidase.
 CC
 SQ Sequence 3108 BP; 888 A; 760 C; 714 G; 746 T; 0 other;
 SX
 QY 1 TTTCAGTGGCTACTACATTTGGCGTGGAGAGCGGATGTCACAGCAACTGTACAGATATC 60
 DB 948 TTTCAGTGGCTACTACATTTGGCGTGGAGAGCGGATGTCACAGCAACTGTACAGATATC 1007
 QY 61 AATTGATGCGCTATGCGCAAAATGCGCAATGCGAGGCTCTCTCAACGAGCTGTC 120
 DB 1008 AATTGATGCGCTATGCGCAAAATGCGCAATGCGAGGCTCTCTCAACGAGCTGTC 1067
 QY 121 AGCCGCTTTATATCTTGGCGGATCCAGATGGGTCAAAATGCAATGGCATTTTGAAGCGTG 180
 DB 1068 AGCCGCTTTATATCTTGGCGGATCCAGATGGGTCAAAATGCAATGGCATTTTGAAGCGTG 1127
 QY 181 GAACATATGATATTTCCCAAGAGCTGAGGTGGAAGTCTCTGAAGAGCTATAGAGATAA 240
 DB 1128 GAACATATGATATTTCCCAAGAGCTGAGGTGGAAGTCTCTGAAGAGCTATAGAGATAA 1187
 QY 241 TATGCTCTCTATATGAGAGACAAATGTTATCTGAGTGCATTCACACACACTCTGCGC 300
 DB 1188 TATGCTCTCTATATGAGAGACAAATGTTATCTGAGTGCATTCACACACACTCTGCGC 1247
 QY 301 CCAGCAGGTTTTTCCATATACACTCTATATACCTGCGCAGGAGGATTCAGCAACCGG 360

Db 1248 CCAGCAGGGTTTTCATATATACACTATATACCTGCCAGCGAGGATTCAGCAACCGG 1307
QY 361 ACCTTTGATAGATAGTCTCTGGGATCATGAAGAGCATTTGATATAGCTCACCAATCTT 420
Db 1308 ACCTTTATAGTATAGTCTCTGGGATCATGAAGAGCATTTGATATAGCTCACCAATCTT 1367
QY 421 AAACGAGGCAAAATCTTATCAACAAAGAAATGTTCTATATGTCAGATCAACCGAAGC 480
Db 1368 AAACGAGGCAAAATCTTATCAACAAAGAAATGTTCTATATGTCAGATCAACCGAAGC 1427
QY 481 CCCCTCTTACCTTCTGGAATCCAGATGAGAGAGAGATTTCTTAAACAGAGC 540
Db 1428 CCCCTCTTACCTTCTGGAATCCAGATGAGAGAGAGATTTCTTAAACAGAGC 1487
QY 541 AAGGAATGCTGGCTTGAATGCTGATTTGATGAGAGAGATTTGCTTATAGC 600
Db 1488 AAGGAATGCTGGCTTGAATGCTGATTTGATGAGAGAGATTTGCTTATAGC 1547
QY 601 TGGTTTGCATCCACCCCGTGAAGATGACAAATGCAACACTTTGATATAGTACAT 660
Db 1548 TGGTTTGCATCCACCCCGTGAAGATGACAAATGCAACACTTTGATATAGTACAT 1607
QY 661 ATGGGCTATGGGCTTACCTTTTGGAGCAAGAAAGAAAGCAAGGCTATGCTGAGCAG 720
Db 1608 ATGGGCTATGGGCTTACCTTTTGGAGCAAGAAAGAAAGCAAGGCTATGCTGAGCAG 1667
QY 721 GGACCGTTTATGAGAGGCTTGTCTCATCAATCTGAGAGAGCTGATCCCAATCTT 780
Db 1668 GGACCGTTTATGAGAGGCTTGTCTCATCAATCTGAGAGAGCTGATCCCAATCTT 1727
QY 781 GGCCCGCATTTGTCAACACAGAGGAGTCTTGTGACACAGCAAGACACTGTCACAC 840
Db 1728 GGCCCGCATTTGTCAACACAGAGGAGTCTTGTGACACAGCAAGACACTGTCACAC 1787
QY 841 GGTGGGCTTATGATGTCATGATGCGGAGCTGAGACAGATGTTTGAAGACACAC 900
Db 1788 GGTGGGCTTATGATGTCATGATGCGGAGCTGAGACAGATGTTTGAAGACACAC 1847
QY 901 ATTATGAGAGGATCATCTATCAAGAGGCAAGAGAGTGTATGCTCTGCTCCAGAG 960
Db 1848 ATTATGAGAGGATCATCTATCAAGAGGCAAGAGAGTGTATGCTCTGCTCCAGAG 1907
QY 961 GTGACGCGCCAGTGTGAGAGTCAACAGTGGTGTGAACATGACAGATGTGAGGCTCAG 1020
Db 1908 GTGACGCGCCAGTGTGAGAGTCAACAGTGGTGTGAACATGACAGATGTGAGGCTCAG 1967
QY 1021 CTCAATGCCACACACAGTGAAGAGCTGTAAACCTGCTGGGCTAAGATTTGGCGCA 1080
Db 1968 CTCAATGCCACACACAGTGAAGAGCTGTAAACCTGCTGGGCTAAGATTTGGCGCA 2027
QY 1081 GGCAATTTGATGAGTTCGGGCTCAATATTTACAGAGGAAATACGGAAGGGATCCA 1140
Db 2028 GGCAATTTGATGAGTTCGGGCTCAATATTTACAGAGGAAATACGGAAGGGATCCA 2087
QY 1141 TTCTGGGACACTTTTCGGGACAGCTCTTGGGAAACATCTGAAGAATTTGAGAGTGT 1200
Db 2088 TTCTGGGACACTTTTCGGGACAGCTCTTGGGAAACATCTGAAGAATTTGAGAGTGT 2147
QY 1201 CAGAAACCCAAACCAATCTGCTTCAAGTGAAGAGCTGAACATACCAATCTTGGCAA 1260
Db 2148 CAGAAACCCAAACCAATCTGCTTCAAGTGAAGAGCTGAACATACCAATCTTGGCAA 2207
QY 1261 CCAGATTTGATGATGTTGATGATTTGATGATTTGATGATTTGATGATTTGATGATTT 1320
Db 2208 CCAGATTTGATGATGTTGATGATTTGATGATTTGATGATTTGATGATTTGATGATTT 2267
QY 1321 GGGGAATTAACAACCATGTCGGGAGCAAGATTTTCTGAGGCAATTAAGAAATTTGCA 1380
Db 2268 GGGGAATTAACAACCATGTCGGGAGCAAGATTTTCTGAGGCAATTAAGAAATTTGCA 2327
QY 1381 CTTTATGAGATGAAGGATATGACGTTGTTATGCGAGGCTTAAGCAATGTTTATACAT 1440
Db 2328 CTTTATGAGATGAAGGATATGACGTTGTTATGCGAGGCTTAAGCAATGTTTATACAT 2387

QY 1441 TACATTAACATATGAAGATATACAGGCTCAGCGGTACGAGGAGCATCTACAACTAT 1500
Db 2388 TACATTAACATATGAAGATATACAGGCTCAGCGGTACGAGGAGCATCTACAACTAT 2447
QY 1501 GGACCAACACCCCTGTCTGATATCATCACTCTTCAAGAGACCTTGTAAAGCAATTTGCT 1560
Db 2448 GGACCAACACCCCTGTCTGATATCATCACTCTTCAAGAGACCTTGTAAAGCAATTTGCT 2507
QY 1561 ACGGACACAGTGAAGCAATGAGAGAGTGTCCGAGGCTCATTTCTTCAAAAATCTATA 1620
Db 2508 ACGGACACAGTGAAGCAATGAGAGAGTGTCCGAGGCTCATTTCTTCAAAAATCTATA 2567
QY 1621 GCTTACCTTATTTCTAATATTTGCGATAGACCAATTTGGCAAAATTTTGGGAGTGC 1680
Db 2568 GCTTACCTTATTTCTAATATTTGCGATAGACCAATTTGGCAAAATTTTGGGAGTGC 2627
QY 1681 TTGAGGCGAGCAAAACCTGAATACAGAGTGGAGAAAGTGTGAAGTTATATTTGTAGGC 1740
Db 2628 TTGAGGCGAGCAAAACCTGAATACAGAGTGGAGAAAGTGTGAAGTTATATTTGTAGGC 2687
QY 1741 GCTAACCCAAAGATTTAGCAGAGACCAAGACCATCAATCAATCTTCTCACTGTGAGAAA 1800
Db 2688 GCTAACCCAAAGATTTAGCAGAGACCAAGACCATCAATCAATCTTCTCACTGTGAGAAA 2747
QY 1801 TAGGAGACTCTGTAGCTGACTGCGAGATATATGTATACGATGCTGCTGGGAGAGAGG 1860
Db 2748 TAGGAGACTCTGTAGCTGACTGCGAGATATATGTATACGATGCTGCTGGGAGAGAGG 2807
QY 1861 TTTTATTTGGCAAAAGAAATCTGGTCTGAGCAATGCAACAAATATATCTGGCATATTTCCA 1920
Db 2808 TTTTATTTGGCAAAAGAAATCTGGTCTGAGCAATGCAACAAATATATCTGGCATATTTCCA 2867
QY 1921 GATATGCTCTTACCTTGAATCTACAGAAATGAATATTTTGGACCAATTCGGAAGCAGGAA 1980
Db 2868 GATATGCTCTTACCTTGAATCTACAGAAATGAATATTTTGGACCAATTCGGAAGCAGGAA 2927
QY 1981 CTTCGAAACCCGCTGATCATCTGATGCAATTTGAAGAAATTTCTCTTTTGAAGTGTG 2040
Db 2928 CTTCGAAACCCGCTGATCATCTGATGCAATTTGAAGAAATTTCTCTTTTGAAGTGTG 2987
QY 2041 ACTACTTAG 2049
Db 2988 ACTACTTAG 2996

RESULT 4
AAA97635
ID AAA97635 standard; DNA; 4835 BP.
XX
AC AAA97635;
XX
DT 02-FEB-2001 (first entry)
XX
DE Mouse neutral/alkaline ceramidase clone, SEQ ID NO:16.
XX
XX
KW Neutral/alkaline ceramidase; mouse; murine; recombinant production;
KW cellular ceramide content control; antibody; membrane synthesis;
KW lipid engineering; ceramide metabolism; drug development; ds.
OS
OS Mus sp.
OS Synthetic.
XX
XX WO200058448-A1.
XX
XX 05-OCT-2000.
XX
XX 24-MAR-2000; 2000WO-JP01802.
XX
XX 26-MAR-1999; 99JP-0084743.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX

PI Ito M;
XX
DR WPI: 2000-619079/59.
XX Mammalian neutral/alkaline ceramidase applicable as lipid engineering
PT reagent for studying structure and functions of ceramide as well as
PT developing drugs for diseases associated with ceramide metabolism
XX
XX Example 6; Page 67-70; 76pp; Japanese.
XX
CC The invention relates to a mouse neutral/alkaline ceramidase and to
CC nucleic acids encoding it. The invention also relates to expression
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
CC ceramidase, the recombinant production of the ceramidase, an antibody
CC against the ceramidase, and a method for controlling the ceramide
CC content in cells and/or tissues by introducing the gene or its antisense
CC nucleic acid into the cells and/or tissues. The ceramidase can be used
CC as a lipid engineering reagent for studying structure and functions of
CC ceramide as well as developing drugs for diseases associated with
CC ceramide metabolism. The present sequence represents a clone encoding
CC mouse neutral/alkaline ceramidase.
CC
SQ Sequence 4835 BP; 1460 A; 996 C; 1048 G; 1331 T; 0 other:

Query Match 99.9%; Score 2047.4; DB 21; Length 4835;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2048; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TTCAGTGGCTTACTACATTGGCGTGGAGAGCGGATTGGACAGACAGTGTCAATATC 60
DB 663 TTCAGTGGCTTACTACATTGGCGTGGAGAGCGGATTGGACAGACAGTGTCAATATC 722
OY 61 AATTTGATGGCTATGCAAAAATGGCCGAATGCAAGGGGTCTCTCCACAGCGCTGTTT 120
DB 723 AATTTGATGGCTATGCAAAAATGGCCGAATGCAAGGGGTCTCTCCACAGCGCTGTTT 782
OY 121 ACCCGTCTTTATCTCTGGCGGATCCAGATGGGTCAAAATGCAATGGCATTTGTGACGCTG 180
DB 783 ACCCGTCTTTATCTCTGGCGGATCCAGATGGGTCAAAATGCAATGGCATTTGTGACGCTG 842
OY 181 GAATATGATGATTTCCCAAGACAGTGAAGTGGAGTCTTGAAGACTAGAGATAA 240
DB 843 GAATATGATGATTTCCCAAGACAGTGAAGTGGAGTCTTGAAGACTAGAGATAA 902
OY 241 TATGGCTCTCTATGCAAGACAAATGTTATCTGAGTGCATTCACACACACTGTGC 300
DB 903 TATGGCTCTCTATGCAAGACAAATGTTATCTGAGTGCATTCACACACACTGTGC 962
OY 301 CCAGCAGGTTTTTCCAAATATACCTATATCTGCGCAGGAGGATTTCAGCAACCGG 360
DB 963 CCAGCAGGTTTTTCCAAATATACCTATATCTGCGCAGGAGGATTTCAGCAACCGG 1022
OY 361 ACCTTTACATATAGTCTCTGGATCATGAAGACATGATATAGCTACACAAATCTT 420
DB 1023 ACCTTTACATATAGTCTCTGGATCATGAAGACATGATATAGCTACACAAATCTT 1082
OY 421 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTCTAATGTGCAGATCAACCGAAGC 480
DB 1083 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTCTAATGTGCAGATCAACCGAAGC 1142
OY 481 CCGTCTCTTACTCTTGAATCCACAGTACAGAGAGCAAGATATTTCTTCAACACAGAC 540
DB 1143 CCGTCTCTTACTCTTGAATCCACAGTACAGAGAGCAAGATATTTCTTCAACACAGAC 1202
OY 541 AAGCAAAATGCTGTTGAAATGTTGATGAGAGAACTTGGGCTTATACAC 600
DB 1203 AAGCAAAATGCTGTTGAAATGTTGATGAGAGAACTTGGGCTTATACAC 1262
OY 601 TGGTTGGCCATCCACCCGTGAGATGAACAATAGCAACCACTTGTGAATAGTGAACAT 660
DB 1263 TGGTTGGCCATCCACCCGTGAGATGAACAATAGCAACCACTTGTGAATAGTGAACAT 1322
OY 661 ATGGGCTATGCGGCTTACTTTTGAAGCAAGAAAGACAAAGGCTATGTGCTGAGACG 720

DB 1323 ATGGGCTATGCGGCTTACTTTTGAAGCAAGAAAGAAAGGCTATGTGCTGAGACG 1382
OY 721 GGCAGCTTTGTAGAGAGGCTTTGCTTCATAAATCTCGAGAGCTGTACCAACATTTCT 780
DB 1383 GGCAGCTTTGTAGAGAGGCTTTGCTTCATAAATCTCGAGAGCTGTACCAACATTTCT 1442
OY 781 GGCAGCTTTGTAGAGAGGCTTTGCTTCATAAATCTCGAGAGCTGTACCAACATTTCT 840
DB 1443 GGCAGCTTTGTAGAGAGGCTTTGCTTCATAAATCTCGAGAGCTGTACCAACATTTCT 1502
OY 841 GGTGGGCTTATGATGATGGCCAGGAGCTGTGACAAAGATGTTTGAAGCACACAC 900
DB 1503 GGTGGGCTTATGATGATGGCCAGGAGCTGTGACAAAGATGTTTGAAGCACACAC 1562
OY 901 ATTATGAGAGGATCATCATGAGAAGGCAAGAGCTTATGCTCTCTCCAGAGAG 960
DB 1563 ATTATGAGAGGATCATCATGAGAAGGCAAGAGCTTATGCTCTCTCCAGAGAG 1622
OY 961 GTGACCGGCCAGTGTCTGACCTCACAGTGGTGAACATGACAGATGTGACGCTCAG 1020
DB 1623 GTGACCGGCCAGTGTCTGACCTCACAGTGGTGAACATGACAGATGTGACGCTCAG 1082
OY 1021 CTCATGCGCACACACAGTGAAGAGCTTAACCTGCGCTGACGTTTGCCGCA 1080
DB 1683 CTCATGCGCACACACAGTGAAGAGCTTAACCTGCGCTGACGTTTGCCGCA 1742
OY 1081 GGCACAAATGATGAGATTTGGGGCTCAATATTATACAGAGGAATACGGAAGGGATCCA 1140
DB 1743 GGCACAAATGATGAGATTTGGGGCTCAATATTATACAGAGGAATACGGAAGGGATCCA 1802
OY 1141 TTCTGGGACACTTCTGGGACACAGCTCTTGGGAAACATCTGAAGAGATTAGAGTGT 1200
DB 1803 TTCTGGGACACTTCTGGGACACAGCTCTTGGGAAACATCTGAAGAGATTAGAGTGT 1862
OY 1201 CAGAAACCCAAACCAATCTGCTCAACAGTGAAGCTGACATACCATCTTGGCAA 1260
DB 1863 CAGAAACCCAAACCAATCTGCTCAACAGTGAAGCTGACATACCATCTTGGCAA 1922
OY 1261 CCAGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
DB 1923 CCAGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1982
OY 1321 GGGGAATTAACAACCATGTCGGGAGCAGAGATTTCTGAGGCAATTAAGAAATTTGCA 1380
DB 1983 GGGGAATTAACAACCATGTCGGGAGCAGAGATTTCTGAGGCAATTAAGAAATTTGCA 2042
OY 1381 CTTTATGGATGAAGATATGACCGTGTATGCGAGGCTTAAGCAATGTTATACAT 1440
DB 2043 CTTTATGGATGAAGATATGACCGTGTATGCGAGGCTTAAGCAATGTTATACAT 2102
OY 1441 TACATTTACCAATATGAAGATATGACGCTCAGCGGTACAGGACATCTCAATCTAT 1500
DB 2103 TACATTTACCAATATGAAGATATGACGCTCAGCGGTACAGGACATCTCAATCTAT 2162
OY 1501 GGAACACACACCTCTGTCATATCATCAACCTTTCAGAGACCTTGTAAAGCAATGTCT 1560
DB 2163 GGAACACACACCTCTGTCATATCATCAACCTTTCAGAGACCTTGTAAAGCAATGTCT 2222
OY 1561 ACGGACAGTGAAGGCAATGAGAGTGTGCGGAGCCCTCAATCTTCAAAAATGTAATA 1620
DB 2223 ACGGACAGTGAAGGCAATGAGAGTGTGCGGAGCCCTCAATCTTCAAAAATGTAATA 2282
OY 1621 GCTTCACTTATTTCTAATATTTGCGATAGAGCAACATTTGGGATGATC 1680
DB 2283 GCTTCACTTATTTCTAATATTTGCGATAGAGCAACATTTGGGATGATC 2342
OY 1681 TTGACAGCAGCAAAACCTGAATACAGAGTGGGAGAGTGTGAAGTATATTTGTAGGC 1740
DB 2343 TTGACAGCAGCAAAACCTGAATACAGAGTGGGAGAGTGTGAAGTATATTTGTAGGC 2402
OY 1741 GCTAACCCAAAGAAATTCAGCAGAGAACGACCAATCAAACTTCTCAGCTGTGAGAA 1800


```
Db 1258 CTCATGCCACACACAGTGAAGACGTGAAGCCGCCCTGGGCTACAGTTTCCGCCA 1317
Qy 1081 GGCACAAATGATGAGATTTTCGGGCTCAATATTACAGAGGAACTACGAGAGGGATCCA 1140
Db 1318 GGCACAAATGATGAGATTTTCGGGCTCAATATTACAGAGGAACTACGAGAGGGATCCA 1377
Qy 1141 TTTCGGGACACTCTTCGGGACACAGCTCTTGGGAAAACCTCTGAAAGATTTGAGAGTGT 1200
Db 1378 TTTTGGGACACTCTTCGGGACACAGCTCTTGGGAAAACCTCTGAAAGATTTGAGAGTGT 1437
Qy 1201 CAGAAACCCAAACCAATCTCTGACAGTGAAGAGTGCAGATGACATCTTGGGCAA 1260
Db 1438 CAGAAACCTTAACCAATCTTAATCCACAGGAGAGCTGCAGAAACCTCATCTTGGCAA 1497
Qy 1261 CCAGATATTTGATGATGTTAGATTGTACCGTTGGGCTCTTGGCCATAGCTGATCCCT 1320
Db 1498 CCAGATATTTGATGATGTTAGATTGTACCGTTGGGCTCTTGGCCATAGCTGATCCCT 1557
Qy 1321 GGGGAATTAACCAATCTGCGGACGAAAGATTTTCGAGGCAATTAAGAAAGATTTGCA 1380
Db 1558 GGGGAATTTACCAACCATGTGAGTGCAGAACTTCGTGAGGCAAGTTAAAAAGAAATTTGCA 1617
Qy 1381 CTTTATGGATGAAGATATGACCGTGTATGCGAGGTCTAACCAATGTTTATACAT 1440
Db 1618 CTTTATGGATGAAGATATGACCTGTATGCGAGGCTGACCAATGTTTATACCAT 1677
Qy 1441 TACATTAACCAATATGAAGAATACAGAGCTCAGCGGTACGAGCAGCATCTACATCTAT 1500
Db 1678 TACATTAACCAATATGAAGAATACAGAGCTCAGCGGTATGAGAGCATCTACATCTAT 1737
Qy 1501 GGACCAACACACCTGTGTGCATACATCCAACTCTTCAGAACCTTGTGAAGCAATTTGCT 1560
Db 1738 GGACCAACACACCTGTGTGCATACATCCAACTCTTCAGAACCTTGTGAAGCAATTTGCT 1797
Qy 1561 ACAGACACATAGCAATGACAGATGAGTGTCCGAGCTCCATCTTCAAAAATCTATATA 1620
Db 1798 ACAGATACATAGCAATGACAGATGAGTGTCCGAGCTCTGAGCTCTTCAAAAATCTATATA 1857
Qy 1621 GCTTCACTATTCCTTAATATTCGGGATAGAGACACAAATTTTGGGATGTC 1680
Db 1858 GCTTCACTATTCCTTAATATTCGGGATAGAGACACCAATTTTGGGATGTC 1917
Qy 1681 TTGCGAGCCAGCAAAACCTGAAATACAGAGTGGGAGAACTGGTGAAGTTATTTGTAGGC 1740
Db 1918 CTGACAGCCAGCAAAACCTGAAATACAGAGTGGGAGAACTGGTGAAGTTGTGAGGC 1977
Qy 1741 GCTAACCCAAAGAAATTCAGCAGAGAACCAACCATCAAACTTCTCAGTGTGAGAAA 1800
Db 1978 GCTAACCCAAAGAAATTCAGCAGAGAACCAACCATCAAACTTCTCAGTGTGAGAAA 2037
Qy 1801 TACGAGACTCTGTAGCTGACGTGACAGATATGATATACAGATGCTCTGGAGACGAGG 1860
Db 2038 TACGAGACTCTGTAGCTGACGTGACAGATATGATATACAGATGCTCTGGAGACGAGG 2097
Qy 1861 TTTTATTTGGCAAAAGGATACGTGGTCTGAGCAATGCAACATATTCGCAATTTCCA 1920
Db 2098 TTTTATTTGGCAAAAGGATACGTGGTCTGAGCAATGCAACATATTCGCAATTTCCA 2157
Qy 1921 GATACGTGCTTACCCCTGGAATCTACAGATTAAGATATTTTGGACAACTCGAAGCAGAA 1980
Db 2158 GATACGTGCTTACCCCTGGAATCTACAGATTAAGATATTTTGGACAACTCGAAGCAGAA 2217
Qy 1981 CTTCTGAAACCCGCTGTCTATAGCATATTTGAAGAAATTTCTTCCCTTTGAAGTTGTC 2040
Db 2218 CTTCTGAAACCCGCTGTCTATAGCATATTTGAAGAAATTTCTTCCCTTTGAAGTTGTC 2277
Qy 2041 ACTACT 2046
Db 2278 ACTACT 2283
```

```
ID AAD22668 standard; cDNA: 2507 BP.
XX
AC AAD22668;
XX
DT 26-FEB-2002 (first entry)
XX
DE . Human ceramidase cDNA.
XX
KW Human; ceramidase; nephrotropic; antipsoriatic; antisense therapy;
KW gene therapy; proliferative disorder; cancer; cardiovascular disease;
KW inflammation; neurodegenerative disorder; cytostatic; immunosuppressive;
KW notropic; signal transduction; breast cancer; autoimmune disorder;
KW Alzheimer's disease; growth deficiency; lesion; lupus nephritis;
KW glomerular disease; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 39..2324
FT /tag= a
FT /product= "Human ceramidase protein"
FT /note= "There is an additional stop codon from position
FT 2325-2327"
FT sig_peptide 39..95
FT /*tag= b
FT mat_peptide 96..2321
FT /*tag= c
FT /product= "Mature ceramidase protein"
PN W0200155A10-A2.
XX
PD 02-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US02866.
XX
PR 28-JAN-2000; 2000US-178975P.
XX
(MUSC-) MUSC FOUND RES DEV.
XX
Hannun YA, El Bawab S;
XX
DR WPI: 2002-025687/03.
XX
PI P-PSDB: AAEI3544.
XX
PT Human mitochondrial ceramidase protein and gene, modulation of which is
PT useful for preventing and treating proliferative disorders e.g. cancer,
PT cardiovascular disease, inflammation and neurodegenerative disorders -
XX
PS Claim 1; Fig 3; 113pp; English.
XX
CC The present sequence is a cDNA encoding human mitochondrial ceramidase
CC protein. Ceramidase protein and gene are useful for treating a disease
CC or disorder involving cell over proliferation or sphingolipid signal
CC transduction especially breast cancer, cardiovascular disorder or
CC inflammation. Ceramidase protein and gene are useful for treating
CC disorders involving deficient cell proliferation or growth e.g.
CC neurodegenerative disorders (Alzheimer's disease), growth deficiencies
CC and lesions. Ceramidase protein is also useful for diagnosis of
CC hyperproliferative diseases. Ceramidase gene can be used as an immunogen
CC to generate antibodies which are useful for diagnosis, prevention and
CC treatment of hyperproliferative diseases and for detecting ceramidase
CC gene product in a biological sample. The hyperproliferative disorders
CC include cancers and autoimmune disorders such as lupus nephritis,
CC glomerular disease. Ceramidase gene is also useful in antisense therapy
CC and gene therapy.
XX
SQ Sequence 2507 BP; 720 A; 583 C; 567 G; 637 T; 0 other;
Qy Query Match 73.8%; Score 1511.4; DB 24; Length 2507;
Best Local Similarity 83.6%; Pred. No. 0;
Matches 1713; Conservative 0; Mismatches 336; Indels 0; Gaps 0;
Qy 1 TTTCAGTGGCTACTACATTTGGCGTTGGAGACGGAGATTGCACAGGACAGTGTACATATTC 60
```


Db 45 AACAGGCTCAGCATATGAGAGCAGCATGACAAATTTATGAGCCGACCCATTATCTGCTT 104
 OY 1523 ACATTCACATCTTGAAGACCTTGTATAGCAATTTGACAGACACATAGCAACATG 1582
 Db 105 ACATTCAGCTTTCAGAACCTTGTATAGGCTATTTGACAGACAGGATGACCACTGTA 164
 OY 1583 GCAGTGTGCTCCGAGCTCCATCTTCTCAAAAATCTAATAGTCTCACTTTCTCATATATG 1642
 Db 165 GCAGAGGTCCAGAACCTCCCTTTTCAACAATTAATAGTTCATTAATCTCATATATG 224
 OY 1643 CGATAGACGACCAATTTGGCAAACTTTGGGATGCTTCCAGCCAGCAAAACCTGAT 1702
 Db 225 TGAGTAGAGCACCAGAACCTTTCGGGATGCTCTCCAGCCAGCAAGAAACCTGAT 284
 OY 1703 ACAGATGGGAGGAGTGTGAATATTTAGAGGCGCAACCCAGAAAGATTCAGAG 1762
 Db 285 ACAGATGGGAGGAGTGTGAATATTTAGAGGCGCAACCCAGAAATTCAGATG 344
 OY 1763 AGAACGAGACCCATCAAACTTCTCACTGTGAGAAATAGAGAGACTTGTAGCTGACT 1822
 Db 345 AAAACGAGACCCATCAAGACTTCTCACTGTGAGAAATAGAGGCTAATTCACATGCT 404
 OY 1823 GCGAGATATATATAGAGCTCTCTGAGAGAGAGGCTTTATTTGCAAAAGAGATAC 1882
 Db 405 GCGAGATATATATAGAGCTCTCTGAGAGAGGCTTTATTTGCAAAAGAGACTC 464
 OY 1883 TGGGCTGAGCAATGCAACATATATAGCTGCAATTCAGATGCTTACCTGCAATCT 1942
 Db 465 TGGGCTGAGCAATGCAACATATGAGATGCAATTCAGATGCTTACCTGCAATCT 524
 OY 1943 ACAGAAATAGATATTTTGGACACATCGAAGCAGAGAACTTCTGAACCCCTGTCATAC 2002
 Db 525 ACAGAAATAGATATTTTGGACACATCGAAGCAGAGAACTTCTGAAGCTGCTGTCATAC 584
 OY 2003 TAGCATTTGAAGAAATTTCTCTCTTGAAGTTCACATCTTAC 2049
 Db 585 TTTTATTGGAAGGACCTTCCCGCTTTTGAAGTTCATATTTAG 631

RESULT 9
 AAZ06239
 ID AAZ06239 standard; DNA: 1194 BP.
 AC AAZ06239;
 XX
 DT 30-SEP-1999 (first entry)
 DE
 XX Human secreted protein gene No. 21.
 XX
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 PN WO935158-A1.
 PD 15-JUL-1999.
 XX
 PF 06-JAN-1999; 99WO-US00108.
 XX
 PR 07-JAN-1998; 98US-0070704.
 PR 07-JAN-1998; 98US-0070657.
 PR 07-JAN-1998; 98US-0070658.
 PR 07-JAN-1998; 98US-0070692.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.

XX
 PI Brewer LA, Duan RD, Edner R, Lafleur DW, NI J;
 PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
 XX
 DR WPI: 1999-444190/37.
 DR P-PSDB: AA138406, AA138464, AA138465, AA138466, AA138467, AA138468,
 DR AA138469, AA138470, AA138471.
 XX
 PS New isolated human genes and the secreted polypeptides they encode
 PS Claim 1: Page 160; 227pp; English.
 XX
 CC This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number is given in the descriptor line.
 CC The gene can be used to generate fusion proteins by linking to the gene
 CC to a human immunoglobulin Fc portion (e.g. AA206210) for increasing the
 CC stability of the fused protein as compared to the human protein only.
 CC The invention relates to 36 novel genes and their fragments (nucleic
 CC acid sequences: AA206219-206263; amino acid sequences AA138386-Y38498)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 36
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AA206219 for described uses).
 XX
 SQ Sequence 1194 BP; 358 A; 291 C; 249 G; 296 T; 0 other;
 Query Match 21.1%; Score 431.8; DB 20; Length 1194;
 Best Local Similarity 81.1%; Pred. No. 1.8e-125;
 Matches 502; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
 OY 1 TTCATGCTCTACTACATTTGGCGTGTGGAGAGCGGATTCACAGAGCAAGTGTGATATC 60
 Db 426 TTCATGCTCTACTACATTTGGCGTGTGGAGAGCGGATTCACAGAGCAAGTGTGATATC 485
 OY 61 AATTGATGGGCTATGAGCAAAATATGCGCAGATGACGCGGCTCTCCAGCAGGCTGTTTC 120
 Db 486 AATTGATGGGCTATGAGCAAAATATGCGCAGATGACGCGGCTCTCCAGCAGGCTGTTTC 545
 OY 121 AGCGTGTCTTATCTTGGCGGATTCAGATGGGCTCAATGCAATGCAATTTGTGAGCGTG 180
 Db 546 AGTGTGCTTCATCATGTCGAGCAAGCTGATGGGCTCAATGCAATGCAATTTGTGAGCGTAC 605
 OY 181 GACATATGATATTTCCCAAGACTGAGTGTGAGGCTCTGAGAGACTAGAGCTAAA 240
 Db 606 GACATATGATATTTCCCAAGACTGAGTGTGAGGCTCTGAGAGACTAGAGCTAAA 665
 OY 241 TATGCTCTCTGTATGCAAGAGACAAATGTTATCCGATGCGCATTCACACACACTCTGCG 300
 Db 666 TATGCTCTCTGTATGCAAGAGACAAATGTTATCCGATGCGCATTCACACACACTCTGCG 725
 OY 301 CCAGCAGGCTTTTCCATATATACACTATATATCTGCGCAGCAGAGGATTCAGCAACCGG 360
 Db 726 CCGTGAAGATATTTCCATATATACCGTGTGTTGAATTTGCGAGGAAGATTTAGCAATCAA 785
 OY 361 ACCTTGTGATATGCTCTGCGGATTCATGGAAGACATTTGATATGCTCACCAAACTCTT 420
 Db 786 ACTTTTCAAGACATGTCATGCTGTTCTTGAAGACATTTGATATGCTCACCAAAATATG 845
 OY 421 AAACGAGCAAAATTTTATATCAAAAGAAATGTTGCTAATGTCACATCAACCGAAG 480
 Db 846 AAACGAGCAAAATTTTATATCAAAAGAAATGTTGCTAATGTCACATCAACCGAAGT 905
 OY 481 CCTCTCTTACCTTCTGAATTCACAGTACAGAGAGCAAGGATTTCTTCAACACAGAC 540
 Db 906 CCGTATCTTACCTTCAAAATCCGACAGTACAGAGAGCAAGGATTTCTTCAAAATACAGAC 965
 OY 541 AAGGAATGCTGTCTTGAAGCTGTGATTTGAATGAGAGAACTTGGGCTTATACGC 600
 Db 966 AAGGAATGATAGTTTGAAGAGTGAATTTGAATGAGATGATGAGGCTTATACAGT 1025


```

QY 129 TTTATCTTGGCGGATCCAGATGGTCAATTCGATGGCATTTGTGAGCGTGGAACTATG 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 630 CTTTGGTGGTGCAGATGAAAAAGG---CAACCGAGTGGCCTGTGGAGCGCAGATGGCGG 686
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 189 TATGATTTCCCAACGACTGAGTGGAGTCTGGAAGAGTACAGAGTAAATATGCTC 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 687 AATGATGGGCTACGAGTTGAAGAGGAGGTGATAAAGACTCCAGGCAGTTATGGCAA 746
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 249 TGTGATCCGAAGACAAATTTATCCGATGGCCATTACACACACTCTGGCCAGAGG 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 747 CATCTACCAAAATGATATGTGGCCATCATGTGGACCCACACACGCTGCTTGGCGG 806
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 309 GTTTTTCATATATACACTTATATACCTGCCAGGAGGATTCAGCAACCGGACCTTCA 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 807 ATTCCGATGCATCTGCTATATGACATCTCATTTGAGGCTTGTGCTTACAGAGCTTTGA 866
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 369 GTACATAGTCTGGGATCATGAAGACATTTGATATAGCTCACACAAATCTTAAACGAG 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 867 GGTGATGGCTCAGGGAGCTGATCTGTGTCATCAAAAGGCGAGCAACCTGGTGTATGG 926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 429 CAAATCTTATATCAAAAGAAATGTGCTAATGTGCAGATCAACCGAACCCCTCTC 488
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 927 TCGCATCTTGTCTCCAAAATCTGTCTAATGTTAATCATCATGCTTCCGCCCTCATC 986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 489 TTACTTCTGATTCACAGTCAAGAGAGCAAGTATTTCTCAAAACAGACAGAAAT 548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 987 CTACTTGAGAAATCCCGCGAGAGAGCGTCCCAATACAGACATACGATTAAGACACT 1046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 549 GCTGGCTTGAACCTGGTGAATTTGAATGAGAAAGCTTGGGTATTATCAGCTGTTGC 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1047 GACCCAGCTCGAATTTGTGCACTGGAAAAACAACCTCCCTGGGCTTTTCAAGTGTATGC 1106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 609 CATTCCACCCCGTGGATGACAAATAGCAACCACTTTGTGAATGTGACATATGGCTA 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1107 GGTGCATGCGCCTCATGATGAACATACAGACAGATGCTGACAGCAATGTGGTTA 1166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 669 TCGGCTTACCTTTTGGAGCAAAAAAGCAAAAGGCTATCTGCTGGAGAGGAGCGTT 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1167 TGCGCCCTCTCTCTGAAAAAGAGATACATCCGAAACAAATGCTCCGAAAGGCAAGTT 1226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 729 TGTAGCAGGCTTGTCTCATCAATCTCGAGAGAGTGTCAACCAATCTTGGCCGCA 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1227 CGTGGGTGCTTGTCTCATCAACCTTGGCGATGTGTCCCAATATATATGAGTCCAA 1286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 789 TTTGTCAACACAGGGGAGTCTGTGACAAAGACAGACAGTCTGTCCACAGCTGGCC 848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1287 GTGCTCGATCTCGGCAATGATGATCTACTGACCTCCGTTGCTCCACTGGCCAG-- 1344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 849 TACATGTGCATGGCCAGCGACCTGAGACAAAGACATGTTGAGAGCACACATTTATAG 908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1345 -GGAGATGTGCTTGTCTCCGAGACCGGTAAAGATATGTTGAAAGCACCAGATCTTGGG 1403
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 909 ACGGATCATCTATCAGAAAGGCCAAGAGCTG-----TATGCTCTGTGCTC 953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1404 TCAACGCTTGTGGGATGTGCTCTGTGGAGTGTCAACGAAACAGAGAGAGTCCAGGGC 1463
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 954 CCAGAGAGTACCGGCGCCAGTGTCTGACGTCAACAGTGGGTAACTGTACAGATGTGAG 1013
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1464 ACGGGAGGTCTACGTGAGATGTGAGTTTCAATCCACAGATGTTGTGTACATGTGCCAATCA 1523
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1014 CGTCCAGCTCAATGCCACAC-----ACAGAGTAAAGCGTGTAAACCTGCG 1058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1524 TGGCAGCAGCTAACAATCCGCTGAGCAGAAAGGTGACAAATTAAGGGAGATGTACGCCGGC 1583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1059 CTTGGCTACAGTTTGGCCGACAGCAATTTGATGAGTTTGGGCCCTCAATATTTACAA 1118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1584 CATGGCTATAGCTTGTCTGTATCCACAGTGAACCTGAGACCTTTCAGCTTCGAGCA 1643
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1119 GGAAGCTACGGAAGGGAGTCTTGTGGACACTCTTCCGGAACAGCTCTTGGGAAAC 1178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1644 GGAAGCCACACGAGACAAACCCATTTGTGCGAGAC---TTATCTGCTGTCTCC 1700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 1179 ATCTGAAGAGATTGTAGAGTGTCAAGAAACCCAAACCAATCCTGCTTACAGTGGAGCT 1238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1701 CACGACAGAGACATATCAATGAGCCATGAGCCCAACCCATCTTCTGTGCGACCGGACGAGG 1760
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1239 GACGATACACATCTTGGCAACAGATATTTGTGATGTTACAGATTTTACCGTGGGTC 1298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1761 CACTTTCCTATAGTGGAGCGCCAAAGATTTGTCTGCAATCAGTACTAAAGATCGGCGA 1820
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1299 CTTGGCATATGCTCTATCCCTGGGGAATTAACAAACCATGTCGGGAGCAAGATTTGCTGA 1358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1821 TGTGATATAGCCGCGCTGCGCTGCGAGTTCCACCAATGAGAGAGCTGTGCGAAA 1880
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1359 GGCATATTAAGAAATTTGACATTTATGGAGTAAGATATGACCTGTTATTCGAGG 1418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1881 CCAATTTGAGACACTGCTCCGCGCTGCGAGGATGACACGAGGTATATCATGCGCGG 1940
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1419 TCTAAGCAATGTTTATACATTTACATTTACCAATATGAAGATACAGGCTCAGCGGTA 1478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1941 ACTGACCAATATATACACAGCTACAGGTGACCCCGAGAGATACAGGCGCAGCGTTA 2000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1479 CGAGGACGATCTATCATATGAGACACACACCTGTCTGATACATATCCAACTCTCAG 1538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2001 CGAGGCGCGCTCCACGATCTTGGGTCCCAATCCACTCATTTACATGAGAGCTTTGCA 2060
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1539 AGACTTGTATGAGCAAT 1556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2061 ACGGCTGACCAAGGCCAT 2078
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 12
AAC03574
ID AAC03574 standard; cDNA; 299 BP.
XX
AC AAC03574;
XX
XX 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 3572.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI: 2000-500381/45.
XX
P-PSDB: AAG03568.
XX
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 3572; 71bp + CD-ROM; English.
XX
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been

```

CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.

XX Sequence 299 BP; 84 A; 68 C; 76 G; 68 T; 3 other:

Query Match 8.1%; Score 166.4; DB 21; Length 299;
Best Local Similarity 82.4%; Pred. No. 7.3e-42;
Matches 211; Conservative 3; Mismatches 39; Indels 3; Gaps 2;

QY 1463 ACCAGGCTCAGCGGATACGACGACATCTACATCTAGACACACCTGTGTGCAT 1522
DB 44 AACAGGCTCAGCGGATACGACGACATCTACATCTAGACACACCTGTGTGCAT 103
QY 1523 ACATGCACTCTTCAGAGACCTTG--CTAGAGCAATTTGACGACACGATGCGCAACAT 1580
DB 104 ACATTCAGCTCTTCAGAAACCTTGCGYMAAGCTATTTGCTACGACACGATGCGCAACCT 163
QY 1581 GAGCAGTGTGCTCCGAGCCTCATCTTCAAAATCTAATAGCTTCACTTATTCCTAATAT 1640
DB 164 GAGCAGAGGTCCAGAACTCCCTTTTCAACA-ATTATAGTTCCTAATTCCTAATAT 222
QY 1641 TCGGATAGACGACCAATTTGCAAAACATTTTGGGATGCTTTCACAGCAGCAAAACCTGA 1700
DB 223 TGTGATAGACGACCAAAAGCAAACTTTCGGGATGCTTCGACGACGCAAAACCTGA 282
QY 1701 ATACAGAGTGGAGAA 1716
DB 283 ATACAGAGTGGGGA 298

RESULT 13

ABLO2670
ID ABLO2670 standard; cDNA; 7083 BP.

XX ABLO2670;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 2492.

KW Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI, 2001-656860/75.

DR P-PSDB; ABB58567.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.

PS Claim 1; SEQ ID NO 2492; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 7083 BP; 2048 A; 1520 C; 1610 G; 1905 T; 0 other:

Query Match 7.6%; Score 155.2; DB 23; Length 7083;
Best Local Similarity 50.6%; Pred. No. 1.8e-37;
Matches 640; Conservative 0; Mismatches 528; Indels 98; Gaps 7;

QY 64 TTGATGCGCTATGCGCAAAATGGCGATGACGAGGCGCTCTCCACAGCGGTTCACG 123
DB 3551 TAGATGCGCTATGCGCAAAATGGCGATGACGAGGCGCTCTCCACAGCGGTTCACG 3610
QY 124 CGTGTCTTATCTTGGCGGATCCAGATGCGTCAAAATGCAATGTCATTTGTGACGCGAA 183
DB 3611 CGTGTCTTATCTTGGCGGATCCAGATGCGTCAAAATGCAATGTCATTTGTGACGCGAA 3667
QY 184 CTATGATATTTCCCAACGACTGAGTGGAGTCTCTCAAGAGACTAGAGATAATAT 243
DB 3668 GCCGGAATGATGCGCTACGATGATGAAGAGGAGTGAATAAAGCACTCCAGCGCATAT 3727
QY 244 GGCTCTCTATGCAAGACAAATGTTATCTGAGGCGCATTCACACACCTCGGCGCA 303
DB 3728 GGCACATCTACCAATATATATGTCATGATGTCACACCGACGCGCGTCTCT 3787
QY 304 GCAGGCTTTTCCAAATATACCTATATATATCTGCGCAGGAGATTCAGCAACCGGACC 363
DB 3788 GCGGATCTCTGATGATGCTGCTATGATGATCTCATCTCGGCTTGTGCTGACGAG 3847
QY 364 TTTCAGTACATGCTCTG-----GATCATGAAG 393
DB 3848 TTTCAGTACATGCTCTG-----GATCATGAAG 3907
QY 394 AGCATGATATGCTCACCAAAATCTTAAACGAGCAAA-----GATCATGAAG 432
DB 3908 GGAACGAATATCTATATATATATATATGATGATCAAAAGGCGAGCAACCTGCTG 3967
QY 433 -----ATCTTATCAAAAGAAATGTTGCTAATGTGAGATCAACCGAACCCCT 484
DB 3968 ATGTCGATCTTGTGCTCAAAAGTCTGCTAATGTGAGATCAACCGAACCCCT 4027
QY 485 CCTTTACCTTCTGATCCAGATCAGAGAGCAAGATATCTTCAAAACAGACAAAG 544
DB 4028 CATCTACTTGAAGATCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4087
QY 545 AATGCTGCTTGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 604
DB 4088 CACTGACCCAGCTGCGATTTGTCGACCTGGAACCAACCTCTCGGCGCTTTCAACTGCT 4147
QY 605 TTGCAATCCACCCGCTGAGATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 664
DB 4148 ATGCGGTGATGCAACCTCATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 4207
QY 665 GCTATGCGCTTACCTTTTGAAGCAAGAAAGCAAGAGATATGCTGCTGACAGGAGC 724
DB 4208 GTTATGCGCGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 4267
QY 725 CGTTTGTGACGAGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 784
DB 4268 AGTTGTGCTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGT 4327
QY 785 CGCATTTGTCTCAACAGAGGAGTGTGTGACAAAGCAAGCAAGCAAGCAAGCAAGCAAG 844
DB 4328 CCAAGTGTGATTCGCGCAATGAGTGTGATCTACTGACTCCGCTGCTGCTGCTGCTG 4387
QY 845 GGCCTAGCATGTGCAATGCGCAGGACCTGCGACAAAGCAATGTTTGAAGAGCACACATTA 904

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 08:54:54 ; Search time 5393 Seconds

(without alignments)
11057.231 Million cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttcaagtgctactactg.....ttgaagtgctactactg 2049

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
GenEmbl1.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_com.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rnd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*
```

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2049	100.0	3108	10	AB037111	AB037111 Mus muscu
2	2047.4	99.9	4835	10	AB037181	AB037181 Mus muscu
3	1833	89.5	2456	10	AB057433	AB057433 Rattus no
4	1830	89.3	2283	6	E50433	E50433 Novel prote
5	1509.8	73.7	2507	6	AX207121	AX207121 Sequence
6	1508.2	73.6	2289	9	AF250847	AF250847 Homo sapi
7	1493	72.9	1571	10	BC022604	BC022604 Mus muscu
8	396.4	19.3	539	6	E50435	E50435 Novel prote
9	264.6	12.9	2140	3	DD082513	DD082513 Dictyosteli
10	192.4	9.4	165407	2	AC117816	AC117816 Mus muscu
11	179.8	8.8	165407	2	AC117816	AC117816 Mus muscu
12	179.8	8.8	184101	2	AC113485	AC113485 Mus muscu
13	158	7.7	235	2	HUMDAC12M5	D17062 Human HepG2
14	155.2	7.6	53059	2	AC012788	AC012788 Drosophila
15	155.2	7.6	173844	3	AC008299	AC008299 Drosophila
16	155.2	7.6	238986	3	AE003774	AE003774 Drosophila
17	150.4	7.3	162228	9	AL450382	AL450382 Human DNA
18	150.4	7.3	176895	2	AL513306	AL513306 Human DNA
19	150.4	7.3	176899	9	AL589794	AL589794 Human DNA
20	143.6	7.0	2618	8	AY057506	AY057506 Arabidops
21	135.2	6.6	105838	2	AC105714	AC105714 Rattus no
22	122.8	6.0	2013	1	PAE315932	AJ115932 Pseudomon
23	122.8	6.0	11622	1	AE004519	AE004519 Pseudomon
24	122.2	6.0	110000	2	AC055726_1	Continuation (2 of
25	122.2	6.0	175330	2	AC012131	Continuation (2 of
26	122.2	6.0	187130	2	AL583858	AL583858 Homo sapi
27	121.2	5.9	1941	6	AR162242	AR162242 Sequence
28	121.2	5.9	1941	6	E62982	E62982 Ceramidase
29	121.2	5.9	2013	6	AR162249	AR162249 Sequence
30	121.2	5.9	2013	6	E62989	E62989 Ceramidase
31	121.2	5.9	2744	1	AB028646	AB028646 Pseudomon
32	115.2	5.6	110000	2	AC055726_2	Continuation (3 of
33	105.6	5.2	77758	8	AB016885	AB016885 Arabidops
34	105.4	5.1	110000	2	AC055726_0	Continuation (2 of
35	105.4	5.1	110000	2	AC055726_1	Continuation (2 of
36	100.6	4.9	175330	2	AC012131	Continuation (2 of
37	80	3.9	19770	1	MRC1376	Z95972 Mycobacteri
38	77.6	3.8	565	6	AX094283	AX094283 Sequence
39	75.6	3.7	153794	8	NC12344	AL67009 Neurospor
40	74.2	3.6	19352	1	AE006964	AE006964 Mycobacte
41	72.4	3.5	184101	2	AC113485	AC113485 Mus muscu
42	72	3.5	154561	8	AP002819	AP002819 Oryza sat
43	71	3.5	200	11	DM208105	Z50514 D. melanoga
44	70.2	3.4	104830	8	AC022464	AC022464 Genomic s
45	67.6	3.3	105838	2	AC105714	AC105714 Rattus no

ALIGNMENTS

RESULT 1
AB037111
LOCUS AB037111 3108 bp mRNA linear ROD 08-JUN-2001
DEFINITION Mus musculus LCBASE mRNA for neutral ceramidase, complete cds.
ACCESSION AB037111.1 GI:7576308
VERSION AB037111.1
KEYWORDS neutral ceramidase.
SOURCE Mus musculus liver cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
Tanl,M., Okino,N., Mori,K., Tanigawa,T., Izu,H. and Ito,M.
AUTHORS
TITLE
Molecular cloning of the full-length cDNA encoding mouse neutral
ceramidase. A novel but highly conserved gene family of

Pred. No. is the number of results predicted by chance to have a

Query Match	Best Local Similarity	100.0%	Pred. No. 0;	Matches 2048;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY 1	TTCAAGTGGCTACTACATATTTGGCGTTTGGGAGACGGGATTGCACAGGACAAGTGTCCAGATATC	60						
DB 663	TTCAAGTGGCTACTACATATTTGGCGTTTGGGAGACGACATTTGCACAGGACAAGTGTCCAGATATC	722						
QY 61	AATTGATGGGCTATGCGAAAAATGGCCAGATATCCAGGGGTCTCTCCACCAGGCTTTC	120						
DB 723	AATTGATGGGCTATGCGAAAAATGGCCAGATATCCAGGGGTCTCTCCACCAGGCTTTC	782						
QY 121	AGCGGTGCTTATATCTGGCGGATCCAGATGGGCTCAATGGAATGGGATTTGTAGCGTG	180						
DB 783	AGCGGTGCTTATATCTGGCGGATCCAGATGGGCTCAATGGAATGGGATTTGTAGCGTG	842						
QY 181	GAACATATGATGATTTCCCAACGACATGAGTTGGAGGTCCTGAAGAGACTAGAGATATAA	240						
DB 843	GAACATATGATGATTTCCCAACGACATGAGTTGGAGGTCCTGAAGAGACTAGAGATATAA	902						
QY 241	TATGGGCTCTCTGATCGAAGACATGTTATCTGAGTGGCATTCACACACACTCTGGC	300						
DB 903	TATGGGCTCTCTGATCGAAGACATGTTATCTGAGTGGCATTCACACACACTCTGGC	962						
QY 301	CCACGAGGTTTTTCCAAATATACACTATATCTGCGACGCGAGGGATTGACGACCGG	360						
DB 963	CCACGAGGTTTTTCCAAATATACACTATATCTGCGACGCGAGGGATTGACGACCGG	1022						
QY 361	ACCTTTAGTACATAGTCTCTGGGATCATGAAAGCATTTGATATAGCTCAACAAATCTT	420						
DB 1023	ACCTTTAGTACATAGTCTCTGGGATCATGAAAGCATTTGATATAGCTCAACAAATCTT	1082						
QY 421	AAACGAGGCAAAATCTTTATACAAAGAAATTTGCTAATGTGCGATACACCGAAGC	480						
DB 1083	AAACGAGGCAAAATCTTTATACAAAGAAATTTGCTAATGTGCGATACACCGAAGC	1142						
QY 481	CCGTCCTCTTACCTTGGAAATCCACAGTCGAGAGACCAAGGTATCTTCAACACAGAC	540						
DB 1143	CCGTCCTCTTACCTTGGAAATCCACAGTCGAGAGACCAAGGTATCTTCAACACAGAC	1202						
QY 541	AAGGAAATGCGTCTTGAAGAACTGTGGATTTGAAATGAGAAAGACTGGGCTCTATACG	600						
DB 1203	AAGGAAATGCGTCTTGAAGAACTGTGGATTTGAAATGAGAAAGACTGGGCTCTATACG	1262						
QY 601	TGTTTTGCCATCCACCCGTGAGCATGACAAATGCAACCACTTTTGGAAATAGTGACAAT	660						
DB 1263	TGTTTTGCCATCCACCCGTGAGCATGACAAATGCAACCACTTTTGGAAATAGTGACAAT	1322						
QY 661	ATGGGCAATGGCGCTTACCTTTTGGACCAAGAAAGCAAAAGGCTATCTGCGTGGACAG	720						
DB 720	ATGGGCAATGGCGCTTACCTTTTGGACCAAGAAAGCAAAAGGCTATCTGCGTGGACAG	780						

```

Db 1323 ATGGGCTATGCGGCTTACCTTTTGAAGCAAAAGCAAGGCTATCTGCTGGACAG 1382
OY 721 GGACCGTTGTAGACAGGCTTGTGCTTATCAATCTCGAGAGCTGTACCCCAATCTT 780
Db 1383 GGACCGTTGTAGACAGGCTTGTGCTTATCAATCTCGAGAGCTGTACCCCAATCTT 1442
OY 781 GGGCCGCTTGTGTCAACACAGGAGGCTTGTGACACAGCAAGACACCTGTCCAC 840
Db 1443 GGGCCGCTTGTGTCAACACAGGAGGCTTGTGACACAGCAAGACACCTGTCCAC 1502
OY 841 GGTGGGCTGTAGATGTGATGGCAGGAGCTGTGACAAAGACATGTTTGGAGACACAC 900
Db 1503 GGTGGGCTGTAGATGTGATGGCAGGAGCTGTGACAAAGACATGTTTGGAGACACAC 1562
OY 901 ATTATAGAGAGATCATCTATTCAGAAAGGCAAGAGCTGTATCTGCTGTCCACAGAG 960
Db 1563 ATTATAGAGAGATCATCTATTCAGAAAGGCAAGAGCTGTATCTGCTGTCCACAGAG 1622
OY 961 GTGACGGGCGGCACTGTGACACTCACAGTGGGTGACATGACATGTGAGCGTCCAG 1020
Db 1623 GTGACGGGCGGCACTGTGACACTCACAGTGGGTGACATGACATGTGAGCGTCCAG 1682
OY 1021 CTCATATGCCACACACAGTGAAGAGGTGTAAACCTGCGCTGGGCTACAGTTTGGCGCA 1080
Db 1683 CTCATATGCCACACACAGTGAAGAGGTGTAAACCTGCGCTGGGCTACAGTTTGGCGCA 1742
OY 1081 GGCACAAATTGATGAGATTTGCGGCTCATATATTACACAGGAACTACGGAAGGAGATCCA 1140
Db 1743 GGCACAAATTGATGAGATTTGCGGCTCATATATTACACAGGAACTACGGAAGGAGATCCA 1802
OY 1141 TTCTGGGACACTTTTGGGACAGCTTGTGGGAAAAACATCTGAAGAGATTTGAGAGTGT 1200
Db 1803 TTCTGGGACACTTTTGGGACAGCTTGTGGGAAAAACATCTGAAGAGATTTGAGAGTGT 1862
OY 1201 CAGAAACCCAAACCAATCTCTGTTCACAGTGAAGAGCTGAGTATACATCTCTGGCAA 1260
Db 1863 CAGAAACCCAAACCAATCTCTGTTCACAGTGAAGAGCTGAGTATACATCTCTGGCAA 1922
OY 1261 CCAGATATTTGATGATGTTGATGTTTACCGTTGGGCTCTTGGCCATAGCTGTATCCCT 1320
Db 1923 CCAGATATTTGATGATGTTGATGTTTACCGTTGGGCTCTTGGCCATAGCTGTATCCCT 1982
OY 1321 GGGGATTTAAACAACGATGTCGGGACGAAGTTTCGTGAGGCAATTTAAAAAGATTTGCA 1380
Db 1983 GGGGATTTAAACAACGATGTCGGGACGAAGTTTCGTGAGGCAATTTAAAAAGATTTGCA 2042
OY 1381 CTTTATGGATGAAGATATGACCGTGTGTTATGCGAGTCTTAAGCAATGTTTATACAT 1440
Db 2043 CTTTATGGATGAAGATATGACCGTGTGTTATGCGAGTCTTAAGCAATGTTTATACAT 2102
OY 1441 TACATTTACACATATGAAAGATATGCAAGGCTCAGCGGTACGAGGACATCTATACATCTAT 1500
Db 2103 TACATTTACACATATGAAAGATATGCAAGGCTCAGCGGTACGAGGACATCTATACATCTAT 2162
OY 1501 GGACGACACACCGCTGTGATGATCATCAACCTTTCAGAGACCTTGTGAAGCAATGTCT 1560
Db 2163 GGACGACACACCGCTGTGATGATCATCAACCTTTCAGAGACCTTGTGAAGCAATGTCT 2222
OY 1561 ACGGACAGTAGACCAATGAGCAGTGTGCGGAGCTCCATCTTCTCAAAAATCTAATA 1620
Db 2223 ACGGACAGTAGACCAATGAGCAGTGTGCGGAGCTCCATCTTCTCAAAAATCTAATA 2282
OY 1621 GCTTCACTTATTTCTAATATGCGGATGAGACCAAAATTTGGGAGTGTG 1680
Db 2283 GCTTCACTTATTTCTAATATGCGGATGAGACCAAAATTTGGGAGTGTG 2342
OY 1681 TTGACACCAAGAAACCTGATATGAGAGTGGGAAAGTGGTTAAGTTATTTGAGGC 1740
Db 2343 TTGACACCAAGAAACCTGATATGAGAGTGGGAAAGTGGTTAAGTTATTTGAGGC 2402
OY 1741 GCTTAACCAAAAGATTCAGCAGAGAACAGACCATCAAACTTCTCACTGTGGAGAA 1800
Db 2403 GCTTAACCAAAAGATTCAGCAGAGAACAGACCATCAAACTTCTCACTGTGGAGAA 2462

```

```

OY 1801 TACGAGACTCTGTAGCTGACTGCGAGATATATTAACGATCTCTCGGAGAGAG 1860
Db 2463 TACGAGACTCTGTAGCTGACTGCGAGATATATTAACGATCTCTCGGAGAGAG 2522
OY 1861 TTTTATTTGGCAAAAGAAATACCTGCTGCTGACAAATGCAATATATCTGGCATTTTCCA 1920
Db 2523 TTTTATTTGGCAAAAGAAATACCTGCTGCTGACAAATGCAATATATCTGGCATTTTCCA 2582
OY 1921 GATACCTGCTTACCTGGAATCTACAGATATAGATATTTTGGACACATCGGAAGAGAA 1980
Db 2583 GATACCTGCTTACCTGGAATCTACAGATATAGATATTTTGGACACATCGGAAGAGAA 2642
OY 1981 CTTCTGAAACCGCTGTGATACATGATGATTTGAAAGAAATTTCTCTTTGAAAGTTGTC 2040
Db 2643 CTTCTGAAACCGCTGTGATACATGATGATTTGAAAGAAATTTCTCTTTGAAAGTTGTC 2702
OY 2041 ACTACTTAG 2049
Db 2703 ACTACTTAG 2711

RESULT 3
LOCUS AB057433 2456 bp mRNA linear ROD 12-JUL-2001
DEFINITION Rattus norvegicus PAPKCD mRNA for ceramidase, complete cds.
ACCESSION AB057433
VERSION AB057433.1 GI:14701597
KEYWORDS
SOURCE Rattus norvegicus kidney cDNA to mRNA.
ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE 1 Mitsutake S., Tanji M., Okino N., Mori K., Ichinose S., Omori A.,
            Iida H., Nakamura T. and Ito M.
            Purification, characterization, molecular cloning, and subcellular
            distribution of neutral ceramidase of rat kidney
            J. Biol. Chem. 276 (28), 26249-26259 (2001)
            21336561
AUTHORS Ito M.
TITLE Direct Submission
JOURNAL Submitted (14-MAR-2001) Makoto Ito, Kyushu University, Bioresource
            and Bioenvironmental Sciences; hakozaiki 6-10-1, Higashi-ku, Fukuoka
            812-8581, Japan (E-mail: makoto@bagr.kyushu-u.ac.jp,
            Tel: 81-92-642-2900, Fax: 81-92-642-2900)
FEATURES
     source
       1..2456
         /organism="Rattus norvegicus"
         /db_xref="taxon:10116"
         /tissue_type="kidney"
       1..2456
         /gene="PAPKCD"
       70..2355
         /gene="PAPKCD"
         /codon_start=1
         /product="ceramidase"
         /protein_id="BAB62033.1"
         /db_xref="GI:14701598"
     translation
       "MAKRIFSSLEAFLLFLVWMAITVALTLFVYSGTEBNHKS
       GNHWSTTGPTTQSSPTTQPTTQPTTQPTLPSPQNSGYIGRADCTGVSDNLM
       GYKGNQNAOGLTTRFSRAFLIADDSGNRAAFVSELCMISORLEVLRLQSKY
       GSLYRBDNYILSATFHSHPAGFGYYLYILASGSGNSRTFOYIVGIVKSIDIAHTN
       LKPKYLINKGNVANYOIRSPSYLONPSEBARYSSDDEMVYLKIDNGEDIG
       LISPPAVHPVSNMNSHNLVNSDMKTAALFLPEKRGRTLPQGPVPAFPASNLCDY
       SPNLTGPHCVNTGESCDNDKSTCPSGGSPCMASGEGQDMFESTHILIGVITOKAEI
       HASASQEVGPLYTAHOWNMNDVSVLAKATHTVKCKALGSFAGTIDVSGSNI
       TOGTTGNTLFDLQDLGKPSSEIIEQAKRPILIHGELTKPPHMQPDIVDIOIV
       TLGSLAIAMIPGEFTTMSGRRLREAVKKEFALYGMNDMTVVVLAGLSNVYTHITFEE
       YXOAREYASTIYGPHTLISAYIOLPFAALKALATDVTVAAGSGPPEPKNLIGSLIP
       NIADRAPIGKQEDVDLPAPKPETRVEVEYFVGCANPKNSAENQTHQFLYVERKD

```

SVANQJMHNDASMETREFFYHKGVLGSLNATIHWHIPDLPVRYIRYFGRKROEL
LKPAVILAEFBISSPFIWTT"
BASE COUNT 704 a 612 c 562 g 578 t
ORIGIN

Query Match 89.5%; Score 1833; DB 10; Length 2456;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 1914; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```

OY 1 TTGAGTGGCTCTACATTTGGCGTTGGGAGCGGATGGACAGACAAATGTCAGATATC 60
   |||||||
DB 307 TTGAGTGGCTCTACATTTGGCGTTGGGAGCGGATGGACAGACAAATGTCAGATATC 366
OY 61 AATTGATGGCTATGGCAAAAATGGCCAGATGACGCGGCTCTCTCCGCCAGCTGTC 120
   |||||||
DB 367 AATTGATGGCTATGGCAAAAATGGCCAGATGACGCGGCTCTCTCCGCCAGCTGTC 426
OY 121 AGCGGTGCTTTTATCTTGGCGGATCCAGATGGTCAATGCAATGGCATTTTGGAGCGTG 180
   |||||||
DB 427 AGCGGTGCTTTTATCTTGGCGGATCCAGATGGTCAATGCAATGGCATTTTGGAGCGTG 486
OY 181 GAATATGATGATTTCCCAAGCAGTGGTGGAGGCTCGAGAGACTGAGAGTAA 240
   |||||||
DB 487 GAATATGATGATTTCCCAAGCAGTGGTGGAGGCTCGAGAGACTGAGAGTAA 546
OY 241 TATGGCTCTCTGATTCGAGAGACATGTTATCCGTGAGTGCATTCACACACACTGCGC 300
   |||||||
DB 547 TATGGCTCTCTGATTCGAGAGACATGTTATCCGTGAGTGCATTCACACACACTGCGC 606
OY 301 CCACAGAGGTTTTCATATACACTCTATATCTCGCCAGCGAGGATTCAGCAACCGG 360
   |||||||
DB 607 CCACAGAGATTTTTCATATACACTCTATATCTCGCCAGCGAGGATTCAGCAACCGG 666
OY 361 ACCTTGATGATAGTCTCTGGGATCATGAAAGCATGATATAGTTCACACAAATCTT 420
   |||||||
DB 667 ACCTTGATGATAGTCTCTGGGATCATGAAAGCATGATATAGTTCACACAAATCTT 726
OY 421 AAACAGGCAAAAATCTTATTCACAAAGAAATGTTGCTATGTCAGATACACGGAAT 480
   |||||||
DB 727 AAACAGGCAAAAATCTTATTCACAAAGAAATGTTGCTATGTCAGATACACGGAAT 786
OY 481 CCTCTCTCTTACCTTCTGAAATCCACAGTCAGAGAGCAGATATCTTCAACACAGAC 540
   |||||||
DB 787 CCTCTCTCTTACCTTCTGAAATCCACAGTCAGAGAGCAGATATCTTCAACACAGAC 846
OY 541 AAGGAATGCGGCTCTGAAACTGCTGATTTGAATGAGAAGCTGGGCTTATCTACGC 600
   |||||||
DB 847 AAGGAATGCGGCTCTGAAACTGCTGATTTGAATGAGAAGCTGGGCTTATCTACGC 906
OY 601 TGGTTGCAATCCACCCCGTAGAGATGAACAAATAGCAACACTTGTGCAATAGTACAT 660
   |||||||
DB 907 TGGTTGCAATCCACCCCGTAGAGATGAACAAATAGCAACACTTGTGCAATAGTACAT 966
OY 661 ATGGGCTATGCGGCTTACCTTTTGGACAAAGAAACAAAGGCTATCTGCTGGAGAG 720
   |||||||
DB 967 ATGGGCTATGCGGCTTACCTTTTGGACAAAGAAACAAAGGCTATCTGCTGGAGAG 1026
OY 721 GGACGCTTGTAGAGGCTTGTCTTCAATCAATCTCGAGACGCTGACCCCAATCTT 780
   |||||||
DB 1027 GGACGCTTGTAGAGGCTTGTCTTCAATCAATCTCGAGACGCTGACCCCAATCTT 1086
OY 781 GGCCCGCATTTGTCAACACAGGAGGAGTCTTGTGACAAAGACAGAGACCTGTCACAC 840
   |||||||
DB 1087 GGCCCGCATTTGTCAACACAGGAGGAGTCTTGTGACAAAGACAGAGACCTGTCACAC 1146
OY 841 GGTGGGCTATGATGTCATGTCAGCGAGCGGACTGAGCAAGACATGTTTGAGAGCACAC 900
   |||||||
DB 1147 GGTGGGCTATGATGTCATGTCAGCGAGCGGACTGAGCAAGACATGTTTGAGAGCACAC 1206
OY 901 ATTATGAGAGGATCATCTATCAAGGCAAGAGAGCTATGCTCTGCTGCCAGAGAG 960
   |||||||
DB 1207 ATTATGAGAGGATCATCTATCAAGGCAAGAGAGCTATGCTCTGCTGCCAGAGAA 1266
```

```

OY 961 GTGACCGGCCACAGTCTTGTCAGCTCACCGAGTGGTGAACATGACAGATGTGAGCGTCAG 1020
   |||||||
DB 1267 GTGACCGGCCACAGTCTTGTCAGCTCACCGAGTGGTGAACATGACAGATGTGAGCGTCAG 1326
OY 1021 CTCATATGCGACACACAGTGAAGACGTGTAAACCTGCTCGGCTACAGTTTGGCCGA 1080
   |||||||
DB 1327 CTCATATGCGACACACAGTGAAGACGTGTAAAGCGGCTCGGCTACAGTTTGGCCGA 1386
OY 1081 GGCACAAATGATGAGGTTTGGGCGCTCAATTTACAGAGGGAACCTACGGAAGGGATCCA 1140
   |||||||
DB 1387 GGCACAAATGATGAGGTTTGGGCGCTCAATTTACAGAGGGAACCAAGAAAGGATCTA 1446
OY 1141 TTCTGGGACACTCTTCGGGACAGCTCTTGGGAAACCATCTGAAGAGATTTGAGAGTCT 1200
   |||||||
DB 1447 TTCTGGGACACTCTTCGGGACAGCTCTTGGGAAACCATCTGAAGAAATCATATGAATTC 1506
OY 1201 CAGAAACCCAAACCATCTCTCTTACAGTGGAGAGCTGACGATACCATCTTGGGCA 1260
   |||||||
DB 1507 CAGAAACCCAAACCATCTATCTCCACAGGAGAGCTGACAAACCTCATCTTGGGCA 1566
OY 1261 CCAGATATTTGATGATTCAGATTTGTCGCTGGGCTCTGGGCAATGCTATATCCCT 1320
   |||||||
DB 1567 CCAGATATTTGATGATTCAGATTCGTTCTTGGTTCGTTGGCCATATGCTATATCCCT 1626
OY 1321 GGGGAAATTAACACCATGTCGGGACGAGATTTGCTGAGGCAATTAAGAAATTTGCA 1380
   |||||||
DB 1627 GGGGAAATTAACACCATGTCAGTGCAGAGCTTGTGGAGGAGTATTAAGAAATTTGCA 1686
OY 1381 CTTTATGGGATGAAGATATGACGTTGTATTCGAGGTCTTAAGCATGTTTATACAT 1440
   |||||||
DB 1687 CTTTATGGGATGAAGATATGACGTTGTATTCGAGGCTGAGCAATGTTTATACCAT 1746
OY 1441 TACATATACATATGAAGATATACAGGCTGAGGAGTTCAGGAGCATCTACATATCTAT 1500
   |||||||
DB 1747 TACATATACATATGAAGATATACAGGCTGAGGAGTTCAGGAGCATCTACATATCTAT 1806
OY 1501 GGACACACACCCCTGCTGATACATCCACTCTTCAGAGACTTGTCTAAGCAATTTGCT 1560
   |||||||
DB 1807 GGACACACACCCCTGCTGATACATCCAGTCTTCAGAGCCCTTGTCTAAGCCATTTGCT 1866
OY 1561 ACGGACAGATGACCAACATGACAGAGTGGTCCGAGCTCCATTTCTTCAAAATCTATA 1620
   |||||||
DB 1867 ACGGACAGATGACCAACATGACAGAGCTGGCTCCATTTCTTCAAAATCTATA 1926
OY 1621 GCTTACTATTTCTTAAATTTGGGATGAGACACCAATTTGCAAAATTTTGGGAGTCT 1680
   |||||||
DB 1927 GCTTACTATTTCTTAAATTTGGGATGAGACACCAATTTGCAAAATTTTGGGAGTCT 1986
OY 1681 TTGACAGCGGCAAAACCTGAATTAAGAGTGGAGAAAGTGTGAAGTTATATTTGTAGGC 1740
   |||||||
DB 1987 TTGACAGCGGCAAAACCTGAATTAAGAGTGGAGAAAGTGTGAAGTTATATTTGTAGGC 2046
OY 1741 GCTAACCCAAAGATTCAGCAGAGAACCAACCATCAACCTTCTCATCTGTGGAGAA 1800
   |||||||
DB 2047 GCTAACCCAAAGATTCAGCAGAGAACCAACCATCAACCTTCTCATCTGTGGAGAA 2106
OY 1801 TACGAGACCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
   |||||||
DB 2107 TACGAGACCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2166
OY 1861 TTTTATTTGGCAAAAGAAATCTGGTCTGATGATGATGATGATGATGATGATGATGATGAT 1920
   |||||||
DB 2167 TTTTATTTGGCAAAAGAAATCTGGTCTGATGATGATGATGATGATGATGATGATGATGAT 2226
OY 1921 GATACGCTTACCTGGAATCTACAGATTAAGATATTTTGGACAAATCGGAGAGGAA 1980
   |||||||
DB 2227 GATACGCTTACCTGGAATCTACAGATTAAGATATTTTGGACAAATCGGAGAGGAA 2286
OY 1981 CTTCTGAACCCGCTGTCATCTAGCAATTAAGAAATTTCTCTTTTGAAGTGTCT 2040
   |||||||
DB 2287 CTTCTGAACCCGCTGTCATCTAGCAATTAAGAAATTTCTCTCTTTTGAAGTGTCT 2346
OY 2041 ACTACTTAG 2049
```

```

Db      2347 ACTACTAG 2355
|||||
RESULT 4
E50433
LOCUS   E50433          2283 bp      DNA          linear      PAT 31-JAN-2002
DEFINITION Novel protein.
ACCESSION E50433
VERSION   E50433.1 GI:18633518
KEYWORDS  JP 2001057890-A/1.
SOURCE    Rattus norvegicus.
ORGANISM  Rattus norvegicus.
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE 1 (bases 1 to 2283)
AUTHORS   Omori, A. and Ito, M.
TITLE     Novel protein.
JOURNAL   Patent: JP 2001057890-A 1 06-MAR-2001;
           MITSUBISHI CHEM CORP
COMMENT   OS Rattus norvegicus (rat)
           PN JP 2001057890-A/1
           PD 06-MAR-2001
           PE 23-AUG-1999 JP 1999235218
           PR
           PI AKIRA OMORI, MAKOTO ITO
           PC C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N9/80, PC
           CC C12N15/00,C12N5/00
           FH Key Location/Qualifiers
           FT CDS (1)..(2283).
           source Location/Qualifiers
           1..2283
           /organism="Rattus norvegicus"
           /db_xref="taxon:10116"
BASE COUNT 641 a 584 c 524 g 534 t
ORIGIN
Query Match 89.3%; Score 1830; DB 6; Length 2283;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 1911; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
1 TTCAGTGGCTACTACATTGGCGTTGGAGAGCGGATTCGACAGGACAGTGTACATATC 60
238 TTCATGTGGCTACTACATTGGCTGTGGCGCTGAGCTGACCTGACAGCAAGTGTACATATC 297
61 AATTGATGGCGCTATGGCAAAATGGCCAGATGGACGGGGTCTCTCCACAGGGCTGTC 120
298 AATTGATGGCGCTATGGCAAAATGGCCAGATGGACGGGGTCTCTCCACAGGGCTGTC 357
121 AGCGTGTCTTTATCTTGGCGGATCCAGATGGGTCAATGCAATGGCATTTGTGACGTG 180
358 ACCCGCGCTTTATCTTGGCGGATCCAGATGGGTCAATGCAATGGCATTTGTGACGTG 417
181 GAACATGATGATATTTCCCAACGAGCTGAGGTGGAGGTCTTGAAGAGACTACAGAGTAA 240
418 GAACATGATGATATTTCCCAACGAGCTGAGGTGGAGGTCTTGAAGAGACTACAGAGTAA 477
241 TTGGCTCTCTATCGAAGACAAATGTTATCTGAGTCCATTGACACACACTGTGC 300
478 TTGGCTCTCTATCGAAGACAAATGTTATCTGAGTCCATTGACACACTGTGTGC 537
301 CCAGAGAGGTTTTTCCAAATATACACTTATATCTGCGCAGGAGGATTCAGCAACCG 360
538 CCAGAGAGATTTTCCAAATATACACTTATATCTGCGCAGGAGGATTCAGCAACCG 597
361 ACCTTTCAGTACATAGTCTCTGGGATCGAAGAGCATTTGATATAGCTACACAATCTT 420
598 ACCTTTCAGTACATAGTCTCTGGGATCGAAGAGCATTTGATATAGCTACACAATCTT 657
421 AAACGAGCAAAATGTTATCAACAAGGAATGTTGATAGTGAAGATCAACGCAACG 480
|||||

```

```

Db      658 AAACGGGCAAAAGTCTTATCAACAAGGAATGTTGCTAATGTGCAGATCAACCGCACT 717
481 CCCCTCCTTACCTTCTGAAATCCAGAGTCCAGAGAGAGCAAGGTATTTCTTCAACACAGAC 540
718 CCCCTCCTTACCTTCTGAAATCCAGAGTCCAGAGAGAGCAAGGTATTTCTTCCGACAGGAC 777
541 AAGGAATGCTGTCTCTTGAAGTGTGATTTGAATGGAAGAGACTTGGGTCTTATACAC 600
778 AAGGAATGCTGTCTCTTGAAGTGTGATTTGAATGGAAGAGACTTGGGTCTTATACAC 837
601 TGGTTTGCATCCACCCCGTGAGAGTGAACCAATAGCAACCACTTGTGAATGTGACAT 660
838 TGGTTTGCATCCACCCCGTGAGAGTGAACCAATAGCAACCACTTGTGAATGTGACAT 897
661 ATGGGCTATGCGGCTTACCTTTTGAAGAAAGAAAGCAAGGCTATCTGCTGGAGACG 720
898 ATGGGCTATGCGGCTTACCTTTTGAAGAAAGAAAGCAAGGCTATCTGCTGGAGACG 957
721 GGACCGTTGTGACAGAGCTTTGCTTATCAATATCTGAGAGAGCTGTACCCAACTTCT 780
958 GGACCGTTGTGACAGAGCTTTGCTTATCAATATCTGAGAGAGCTGTACCCAACTTCT 1017
781 GGCCCGCATTTGTCAACACAGAGGAGTCTTGTGAACAGCAAGAGCACTGTCCCAAC 840
1018 GGCCCGCATTTGTCAACACAGAGGAGTCTTGTGAACAGCAAGAGCACTGTCCCAAC 1077
841 GGTGGGCTTATGATGTGATGGCCAGAGGAGCTTGACCAAGACATGTTTGAAGACACACAC 900
1078 GGTGGGCTTATGATGTGATGGCCAGAGGAGTCTTGTGAACAGCAAGAGCACTGTCCCAAC 1137
901 ATTATAGAGCGGATATCTATCAGAAAGGCAAGAGAGCTTATGCTTCTCTCCAGAGAG 960
1138 ATTATAGAGCGGATATCTATCAGAAAGGCAAGAGAGCTTATGCTTCTCTCCAGAGAG 1197
961 GTGACGCGCCAGAGCTTCTGCTTACACAGTGGGAGTGAACATGACAGATGAGAGCTCCAG 1020
1198 GTGACGCGCCAGAGCTTCTGCTTACACAGTGGGAGTGAACATGACAGATGAGAGCTCCAG 1257
1021 CTCATGCGCACACACAGTGAAGAGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
1258 CTCATGCGCACACACAGTGAAGAGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1317
1081 GGCACATTTGATGAGATTTTGGGCTCCTATTTTACACAGGAACTACAGGAAGGATGCA 1140
1318 GGCACATTTGATGAGATTTTGGGCTCCTATTTTACACAGGAACTACAGGAAGGATGCA 1377
1141 TTCTGGACACTTCTGCGGACAGCTCTTGGGAAACCACTGGAAGAGATTTAGAGT 1200
1378 TTCTGGACACTTCTGCGGACAGCTCTTGGGAAACCACTGGAAGAGATTTAGAGT 1437
1201 CAGAAACCCAAACCACTGCTTACAGTGAAGAGCTGAGATACACATCTCTTGGCA 1260
1438 CAGAAACCCAAACCACTTACAGTGAAGAGCTGAGATACACATCTCTTGGCA 1497
1261 CCAGATATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
1498 CCAGATATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1557
1321 GGGGAATTAACCAACATGTCGGGAGCAGAAATTTCTGAGGCAATTAACCAATTTGCA 1380
1558 GGGGAATTAACCAACATGTCGGGAGCAGAAATTTCTGAGGCAATTAACCAATTTGCA 1617
1381 CTTTATGGGATGAAGATATGACCTGTTATTCGACAGCTCTAAGCAATTTTATACAT 1440
1618 CTTTATGGGATGAAGATATGACCTGTTATTCGACAGCTCTAAGCAATTTTATACAT 1677
1441 TACATTACACATATGAAGAAATATACAGGCTCAGGCTAGAGGCACTATCAATCTAT 1500
1678 TACATTACACATATGAAGAAATATACAGGCTCAGGCTAGAGGCACTATCAATCTAT 1737
1501 GGACGACACACCTGCTGCTGATACATCACTCAACTCTTACAGAGACTTGTCAAGCAATGCT 1560
1738 GGACGACACACCTGCTGCTGATACATCACTCAACTCTTACAGAGACTTGTCAAGCAATGCT 1797

```


OY	1561	ACGGACAGATAGGCCAACATATAGCATAGTGTGTCCCGAGCCTCCATTCCTTCAAAAACCTATA	1620
Db	1798	ACGGATACAGTAGCCACATATAGCAGCGCGGTCTTGAGCTCCATTCCTTCAAAAACCTGATA	1857
OY	1621	GCTTCACCTTAATTCCTAATATATTTGCGGATTTAGAGACCCCAATTTGGCAAAACATTTTGGGGATGTC	1680
Db	1858	GGTTACACTTATTCCTAATATATTTGCGGATTTAGAGACCCCAATTTGGCAAAACATTTTGGGGATGTC	1917
OY	1661	TTTGACAGCCAGCAAAACCTCGAATTACAGAGTGGGAGAGTGGTTGAAGTTATATTTTGTAGGC	1740
Db	1918	CTTGAGCCAGCAAAACCTCGAATTACAGAGTGGGAGAGTGGTTGAAGTTATATTTTGTAGGC	1977
OY	1741	GCTAACCCAAAGAAATTTGCGAGAGACAGACAGCCCATCAACCTTCCTACCTGTGGAGAA	1800
Db	1978	GCTAACCCAAAGAAATTTGCGAGAGACAGACAGCCCATCAACCTTCCTACCTGTGGAGAA	2037
OY	1801	TACGAGACCTCTGTAGCTGACTGGCGACATTAATGTATAACGATGGCTCCTGGGAGACGAG	1860
Db	2038	TACGAGACCTCTGTAGCCACATGGCGACATTAATGTATAACGATGGCTCCTGGGAGACGAG	2097
OY	1861	TTTATTTGCGCAAAAGCAATATCTGGGCTGTGAGCAATGCACAAATATACGCGCATATTCCA	1920
Db	2098	TTTATTTGCGCAAAAGGAGTACTGGGCTGTGAGCAATGCACAAATATACATTTGCATATTCCA	2157
OY	1921	GATACGCGCTCCCTCGGAATCTACAGAAATAGATATTTTGGACCAAAATCGGAGACAGAA	1980
Db	2158	GATACGCGCTCCCTCGGAGCTTACAGAAATAGATATTTTGGACCAAAATCGGAGACAGAA	2217
OY	1961	CTTCTGAACCCCGCTGCTACTACAGCAATTTGAAGAAATTTCTCTCTTTTGAAGTTGTC	2040
Db	2218	CTTCTGAAGCCGCTGCTACTACAGCAATTTGAAGAAATTTCTCTCTTTTGAAGTTGTC	2277
OY	2041	ACTACT 2046	
Db	2278	ACTACT 2283	
RESULT 5			
LOCUS	AX207121	2507 bp	DNA linear PAT 30-AUG-2001
DEFINITION	Sequence 1 from Patent W00155410.		
ACCESSION	AX207121		
VERSION	AX207121.1	GI:15394925	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2507)		
TITLE	Hannun,Y.A. and el Bawab,S.		
JOURNAL	Ceramidase compositions and methods based thereon		
FEATURES	Patent: WO 0155410-A 1 02-AUG-2001;		
SOURCE	MUSC Foundation For Research Development (US)		
	Location/Qualifiers		
	1..2507		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	39..2324		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC60194.1"		
	/db_xref="GI:15394926"		
	/translation="MSAITVALSLIFTSQTIENHKDIGHPSTQSPAPQGSTA		
	AORSTATQHSATQSSRATQSPPLRPESLPQNSGTHGVRADCTQGVADINLM		
	GXGSGNAGCIIRLIKSRATIAEPGSRNTTYSVIDIGVSRLEIVNRLQSKY		
	GSLLKRDVILISGITHSGPAGIVQYTVFVYASGFSNITFOHWATILNSIDIAHTN		
	MKPKRIFIKNGVDGVQINRPSYILNPQSESRYSNIDKEMLVIKMDLNDQD		
	LISMFALPIVSMNSNHLVNSDNGVASYLLEKPCNDGYLPQGPFPVAFASNIGDV		
	SLINFPRICTPAGESCNANSTCPIGGSPMK IAKPGQMDHDSFOIIGRAMYORAKEI		
	YASQSEVTEGPIASHOWVMDVTVLWLNSTHAKTCKPALGYSFAAGTIDVCGGLN		
	TQGTQEDDPEDPTIRDOILGKPSRIECHKPRKRIILHGLSLSPHMHDIYDVQIIT		
	TLGSLAITAIGETTTMSGRKLRAVOAFSRHOMQNNVTVISGLCAVYTHITTYEE		
CDS			

BASE COUNT		720 a		582 c		567 g		638 t		
ORIGIN										
<p>Query Match 73.7%; Score 1509.8; DB 6; Length 2507; Best Local Similarity 83.6%; Pred. No. 0; Matches 1712; Conservative 0; Mismatches 337; Indels 0; Gaps 0;</p>										
QY	1	TTCA	TG	GC	CT	CT	CA	TTG	GGCGG	GGAGACGCGGATTTGCACAGACAAAGTCGATATC
DB	276	TTCA	GT	GG	CT	CA	CT	ATTTGGT	GGACGACCTACTCTACAGACAAAGTAGCAGATATC	
QY	61	AATTTG	AT	GG	CT	AT	GG	CAAAAATG	CGCCAGAAATGCACGGGGTCTCTCACCAGGCTGTC	
DB	336	AATTTG	AT	GG	CT	AT	GG	CAAAAATG	CGCCAGAAATGCACGGGGTCTCTCACCAGGCTGTC	
QY	121	AGCC	GT	CT	TTT	AT	CT	TGGCGG	ATTCAGATGGGCTTGTGACGCTG	
DB	396	AGTG	CT	GC	CT	CA	TC	ATGCGAGAACTG	ATGGGCTTCCATCGAACAGTGTGTGCAATC	
QY	181	GA	CT	AT	GT	AT	TTCC	CA	AGCACTAGAGTTGAGAGTCCGAGAGCTGAGACGTGAGACGTAA	
DB	456	GAC	AT	AG	CA	AT	GG	CA	ATGATATCAAAAGGCTCAGGCTGAGAGTCTCGAACAAGCTGCAAGATAA	
QY	241	TAT	AG	CT	CT	CT	GT	AT	TCGAGAGACAATGTTATCTGAGTGGCATTTCACACACTGTGGC	
DB	516	TAT	GG	CT	CT	CT	GT	AT	TCGAGAGACAATGTTATCTGAGTGGCATTTCACACACTGTGGC	
QY	301	CC	AG	CG	GG	TT	TT	CC	AAATATATACACTCTATATCTGCGCAGCGAGGGATTTCAGACACCGG	
DB	576	CC	TC	AG	CA	AT	AT	TTCC	AGTATATCCGATATACCGGTGTTGTATTTCCACATGGAAGATTATGACAACTAA	
QY	361	AC	CT	TT	CA	GT	AT	AG	CT	ATATGAGAGCAATGATATGATATAGCTCAACAAATCTT
DB	636	AC	TT	TT	CA	GT	AT	AG	CT	ATATGAGAGCAATGATATGATATAGCTCAACAAATCTT
QY	421	AA	AC	CA	GA	CA	AA	AT	CT	TTATCAACAAAGGAATTTGCTATGTGCAGATCAACCGAGC
DB	696	AA	AC	CA	GA	CA	AA	AT	CT	TTATCAACAAAGGAATTTGCTATGTGCAGATCAACCGAGC
QY	481	CC	GT	CT	CT	TA	CT	GA	AT	TCACACAGCTCAGAGAGCAAGATATCTTCAACACAGAC
DB	756	CC	GT	AT	CT	CT	TA	CT	GA	AT
QY	541	A	A	G	A	A	A	T	G	C
DB	816	A	A	G	A	A	A	T	G	C
QY	601	T	G	T	T	G	C	A	T	T
DB	876	T	G	T	T	G	C	A	T	T
QY	661	A	T	G	G	C	T	A	T	G
DB	936	A	T	G	G	C	T	A	T	G
QY	721	G	G	A	C	G	T	T	G	A
DB	996	G	G	A	C	G	T	T	G	A
QY	781	G	G	C	C	G	A	T	T	G
DB	1056	G	G	C	C	G	A	T	T	G
QY	841	G	G	T	G	G	C	T	A	T
DB	1116	G	G	T	G	G	C	T	A	T
QY	901	A	T	T	A	G	A	C	G	A
DB	960	A	T	T	A	G	A	C	G	A

Db 1176 ATTATAGAGGGCCATATATACAGAGCAAAAGAACTCTATGCTTGCCTCCAGGAG 1235
 QY 961 GTGACCGGCCAGTGCCTTGCAGCTCACCAGTGGGTGAACATGACAGATGTGAGGCTCCAG 1020
 Db 1236 GTAAAGAGGACACAGCTTCAAGACAGAGTGGGTGATGATGACAGATGTGACCTCTGG 1295
 QY 1021 CTCAATGCGACACACAGCTGTAAGAGCTGTAACTGCCCTGGGCTACAGTTTCCCGCA 1080
 Db 1296 CTCAATTCACACATGATCAAAAACATGTAAACAGATGTGGGCTACAGTTTCCGCT 1355
 QY 1081 GGCAATTTGATGAGTTTGGGCTCCTCAATATTACAGGGGAACACAGGAAGGATGCA 1140
 Db 1356 GGCATATTGATGAGTTGAGGCTCCTCAATTTTACACAGGGGAACAGGAAGGATGCA 1415
 QY 1141 TTCTGGGACACTTCTGGGACAGCTCTTGGGAAAACCATCTGAAGAGATTGTAGAGT 1200
 Db 1416 TTTTGGGACACCATCTGGGACAGATCTGGGAAAACCATCTGAAGAGATTGAAGATGT 1475
 QY 1201 CAGAAACCCAAACCAATCTGCTTACAGTGAAGAGCTGACAGATACCATCTTGGCAA 1260
 Db 1476 CATAAACCAAAAGCCCTCTTCTTACACCGGAGAACTATCAAAACCTCACCCCTGGCAT 1535
 QY 1261 CCAGATATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 Db 1536 CCAGATATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1595
 QY 1321 GGGGATTTAGAACCATCTGCGGACGAGAAATTTCTGAGGCAATTAAGAAATTTGCA 1380
 Db 1596 GGGGATTTAGAACCATCTGCGGACGAGAAATTTCTGAGGCAATTAAGAAATTTGCA 1655
 QY 1381 CTTATGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 Db 1656 TCTCTGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1715
 QY 1441 TACATTTACCATATGAGAAATACAGGCTCAGCGTACGAGAGCATCTACATCTAT 1500
 Db 1716 TACATTTACCATATGAGAAATACAGGCTCAGCGTACGAGAGCATCTACATCTAT 1775
 QY 1501 GGACACACACCCCTGCTGATACATTCACACTCTTCAAGAGACCTTCTGAGGCAATTTGCT 1560
 Db 1776 GGACACACACCCCTGCTGATACATTCACACTCTTCAAGAGACCTTCTGAGGCAATTTGCT 1835
 QY 1561 AGGACACAGTACGCAATGATGAGCTGCTCCGAGCTCCCTCTTCAAAATATTAATTA 1620
 Db 1836 AGGACACAGTACGCAATGATGAGCTGCTCCGAGCTCCCTCTTCAAAATATTAATTA 1895
 QY 1621 GCTTCACTTATTTCTAATATTTGCGGATAGAGACCAATTTGCGGATGATG 1680
 Db 1896 GCTTCACTTATTTCTAATATTTGCGGATAGAGACCAATTTGCGGATGATG 1955
 QY 1681 TTGCAGCCAGCAAAACCTGAATATACAGATGGGAGAGTGTGAAGTTATATTTAGGC 1740
 Db 1956 CTGCAGCCAGCAAAACCTGAATATACAGATGGGAGAGTGTGAAGTTATATTTAGGC 2015
 QY 1741 GCTAACCCCAAAAGAAATTCAGAGAGAGAACCAACCTCAACCTTCTCAGCTGAGGAAA 1800
 Db 2016 GCTAACCCCAAAAGAAATTCAGAGAGAGAACCAACCTCAACCTTCTCAGCTGAGGAAA 2075
 QY 1801 TACAGAGACTCTGTAGTACGAGTGTGAGAGATTAATGATTAAGATGCTCCTCGGAGAGAG 1860
 Db 2076 TATAGAGCTACTTCAACATCGTGTGAGAGATTAATGATTAAGATGCTCCTCGGAGAGAG 2135
 QY 1861 TTTTATTTGGCAAAAGAAATCTGAGTGTGAGAGCAATGCAAAATATCTAGGCAATTTCCA 1920
 Db 2136 TTTTATTTGGCAAAAGAAATCTGAGTGTGAGAGCAATGCAAAATATCTAGGCAATTTCCA 2195
 QY 1921 GATAGTGGCTACCTGGAATCTAGAGATTAATTTGGAGACAAATGCGAAGAGGAA 1980
 Db 2196 GACATGCGCCAGCCCTGGAATCTAGAGATTAATTTGGAGACAAATGCGAAGAGGAA 2255
 QY 1981 CTTTCAACCCGCTGTGATACATTTGAGAGAAATTTCTCTTTGAGAGTGTG 2040
 Db 2256 ATTCTGAAGCTGCTGTACTACTTCAATTTGAGAGCACTTCCCGGCTTTTGAAGTTGTA 2315

QY 2041 ACTACTAG 2049
 Db 2316 ACTATTAG 2324
 RESULT 6
 AF250847
 LOCUS
 DEFINITION
 Homo sapiens mitochondrial ceramidase mRNA, complete cds; nuclear gene for mitochondrial product.
 AF250847
 VERSION
 AF250847.1 GI:9246992
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 2289)
 El Bawab, S., Roddy, P., Qian, T., Bielawska, A., Lemasters, J.J. and Hannun, Y.A.
 Molecular cloning and characterization of a human mitochondrial ceramidase
 J. Biol. Chem. 275 (28), 21508-21513 (2000)
 MEDLINE
 PUBMED
 10781606
 2 (bases 1 to 2289)
 El Bawab, S. and Hannun, Y.A.
 Direct Submission
 Submitted (30-MAR-2000) Biochemistry, Medical University of South Carolina, 114 Doughty Street, P.O. Box 250780, Charleston, SC 29425, USA
 FEATURES
 source
 1..2289
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..2286
 /function="cleaves lipid ceramide into free fatty acid and sphingosine"
 /note="active in neutral to alkaline range"
 /codon_start=1
 /product="mitochondrial ceramidase"
 /protein_id="AF250847.1"
 /db_xref="GI:9246993"
 /translation="MSATVVALSLFTTSGTIEHKDLGHEFTSTGSPATGCTA
 AQRSTATSHSTATOSTATQSPVLPETSPFQNFSGYHIGVGRADCTGVADINLM
 GYKSGQNAOGLITRLYSRAFLMAEPDSNRTVEYSIDIGVSOPLRIEVLNLSQSY
 GLVSRDNLVILSGTHSGPAGYFQYTFVFLASBEFSNOTOHVYTGILKSLADHTN
 MKPKRTINKGNQVQVQINRSYSYLOMPOSERAYPSTNTKEMVILKMDLNDLIG
 LLSWFAIHPVSKNNNSNHLVSDNVGTASYLLEQENKGYLPGQFFVAIFSSNLGDV
 SPNLIQPRICINTGESCDNANSTCPDIPGSPSKTAKPGQDMFDSTQIIGRAMYORAKEL
 YASAEVETGPLASHQWMDMDVYVWLNSTHASKTCQPALGYSFAACTIDGVGLNF
 TQSGQEPFMDITRIDQILKPSSEIKKCHPKPILTLHGTGLSKHPWHPDIDVQIL
 TLGSLAIFAIQGEFTTMSGRRLREAVQEPVLSHGQONTVVISGLCNYTHYITTYEE
 YQAKRYEASRTYGPALASAYIQFLRNALADDTVNLISRGPEPKQIYDPLIE
 STVDRAPEKRTFQDVLQPAKPEIRYGEVAEYIVGANKNSYQNDHOTFLVEKTEA
 TSTSWQIVCNDASWETREYVWHLGGLISLNATVEHMLPDTAPGILYRIRYFQHNRKQDI
 LKPAVILSEFTSPAFEVYTI"
 ORIGIN
 664 a 553 c 515 g 557 t
 Query Match 73.6%; Score 1508.2; DB 9; Length 2289;
 Best Local Similarity 83.5%; Pred. No. 0;
 Matches 1711; Conservative 0; Mismatches 338; Indels 0; Gaps 0;
 QY 1 TTCAGTGGCTACTACTATGGGCTTGGGAGAGCGGATTCACAGCAAGTGTACAGATATC 60
 Db 238 TTCAGTGGCTACTACTATGGGCTTGGGAGAGCGGATTCACAGCAAGTGTACAGATATC 297
 QY 61 AATTGATGGGCTATGGCAAAATGCGCAGATGCGAGGGGTCTCTCACCAGGCTGTTC 120
 Db 298 AATTGATGGGCTATGGCAAAATGCGCAGATGCGAGGGGTCTCTCACCAGGCTATAC 357

QY 121 AGCGTGCTTTATCTTGGGGATCAGATGGGTCAATGCAATGGCATTTGTGGCGTG 180
 DB 358 AGTGTGCTTATCATTTGATGGGAGAACCTGATGGTCCATGCAAGTGTGTGACATC 417
 QY 181 GAACTATGATGATTTCCCAACAGCTAGTGGAGTCTTGAAGAGCTAGAGATAA 240
 DB 418 GACTATGAGCATGATGATCAAAAGGCTCAGGCTGAGGCTCCGAAACACAGTCAAGTAA 477
 QY 241 TATGGCTCTGATGATGAGAGCAATGATATCTGTGATGCTATCCATCAACACTGCGC 300
 DB 478 TATGGCTCTGATGATGAGAGCAATGATATCTGTGATGCTATCCATCAACACTGCGC 537
 QY 301 CCAGCAGGTTTTTCCATATACACTATATCTCCAGCGAGGATTCAGACACCG 360
 DB 538 CTGCGAGATATTTCCAGTATACCGTGTGTATTTGATGCTGAGGATTTAGCATCA 597
 QY 361 ACCTTCAGTACATAGTCTCTGGGATCAGAGCAATGATATGCTCAGCAAACTTT 420
 DB 598 ACTTTTCAGCATAGTCTCTGGGATCAGAGCAATGATATGCTCAGCAAACTTTATG 657
 QY 421 AAACGAGCAAAATCTTATCAACAAAGAAATGCTATGTCAGATGCAACGAGC 480
 DB 658 AAACGAGCAAAATCTTATCAACAAAGAAATGCTATGTCAGATGCAACGAGC 717
 QY 481 CCTCTCTTACCTTCTGAATCCACAGTCAAGAGACAGATATCTTCAACACAGAC 540
 DB 718 CCGTATCTTACCTTCAAAATCCGAGTCAAGAGACAGATATCTTCAAAATACAGAC 777
 QY 541 AAGAAATGCTGGCTTGAAGAGTGGTGAATTTGAATGAGAGCACTGGGCTTATCAGC 600
 DB 778 AAGAAATGATAGTTTGAAGAAATGATGATTTGAATGAGATGCTGGGCTTATCAGC 837
 QY 601 TGGTTTGCATCCACCCGTCAGATGACATGACCAATGACCACTTGAATGAGCAAT 660
 DB 838 TGGTTTGCATCCACCCGTCAGATGACATGACCAATGACCACTTGAATGAGCAAT 897
 QY 661 ATGGGCTATGCGGCTTACCTTTTGAAGCAAGAAAGCAAGGCTATGCTGCTGAGAC 720
 DB 898 GTGGGCTATGCGGCTTACCTTTTGAAGCAAGAAAGCAAGGCTATGCTGCTGAGAC 957
 QY 721 GGACCGTTTGTAGAGGCTTGTCTTCAATTCGAGAGCTGTCAACCACTTCTT 780
 DB 958 GGCGCATTTGTAGAGGCTTGTCTTCAATTCGAGAGCTGTCAACCACTTCTT 1017
 QY 781 GGCGCGCATTTGTAGAGGCTTGTCTTCAATTCGAGAGCTGTCAACCACTTCTT 840
 DB 1018 GGACCGATTTGTAGAGGCTTGTCTTCAATTCGAGAGCTGTCAACCACTTCTT 1077
 QY 841 GGTGGGCTATGATGCTATGCTGAGGAGGCTGAGACAGATGTTTGAAGGACACAC 900
 DB 1078 GGTGGGCTATGATGCTATGCTGAGGAGGCTGAGACAGATGTTTGAAGGACACAC 1137
 QY 901 ATATATGAGAGGATGATGATGATGAGAGGCAAGGAGCTGTATGCTCTGCTCCAGAG 960
 DB 1138 ATATATGAGAGGAGGCTGATGATGAGAGGCAAGGAGCTGTATGCTCTGCTCCAGAG 1197
 QY 961 GTGACGCGGAGTGTGCTGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAG 1020
 DB 1198 GTGACGCGGAGTGTGCTGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAG 1257
 QY 1021 CTCATGTCACACACAGATGTAAGAGCTGTAAACCTGCTGCTACAGTTTTCGCGCA 1080
 DB 1258 CTCATGTCACACACAGATGTAAGAGCTGTAAACCTGCTGCTACAGTTTTCGCGCT 1317
 QY 1081 GGCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 DB 1318 GGCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1377
 QY 1141 TTCTGGAGACACTCTTGGGAGCACTTGGGAGAAACATCTGAAGATTTGAGAGTGT 1200
 DB 1378 TTCTGGAGACACTCTTGGGAGCACTTGGGAGAAACATCTGAAGATTTGAGAGTGT 1437
 QY 1201 CAGAAACCAAAACCAATCTGCTTCAAGTGGAGAGCTGACATACCATCTTGGCAA 1260

DB 1438 CATTAACCAAGCCCATCTTCTTACACCCGAGAACATATCAAAACCTCACCCCTGGCAT 1497
 QY 1261 CCAGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 DB 1498 CCAGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1557
 QY 1321 GGGGAATTAACACATGCTGGGAGGAGGATTTGCTGGGAGGATTTAAAGAAATTTGCA 1380
 DB 1558 GGGGAGTTTACGACATGCTGGGAGGAGGATTTGCTGGGAGGATTTAAAGAAATTTGCA 1617
 QY 1381 CTTATGAGGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 DB 1618 TCTATGAGGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1677
 QY 1441 TACATTAACATATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 DB 1678 TACATTAACATATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1737
 QY 1501 GGACCAACACCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
 DB 1738 GGACCAACACCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1797
 QY 1561 ACGGACAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 DB 1798 ACGGACAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1857
 QY 1621 GCTTACATTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 DB 1858 GTTCCATTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1917
 QY 1681 TTTGAGGAGCAAAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 DB 1918 CTGAGGAGCAAAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1977
 QY 1741 GCTAACCCAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
 DB 1978 GCTAACCCAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2037
 QY 1801 TACGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
 DB 2038 TATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2097
 QY 1861 TTTTATGAGCAAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
 DB 2098 TTTTATGAGCAAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2157
 QY 1921 GATACGCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 DB 2158 GATACGCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2217
 QY 1981 CTTTGAAGCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040
 DB 2218 ATTCTGAAGCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2277
 QY 2041 ACTACTTAG 2049
 DB 2278 ACTATTTAG 2286

RESULT 7
 BC022604
 LOCUS
 DEFINITION Mus musculus. Similar to N-acetylserine aminohydrolase 2, clone
 IMAGE:4162513, mRNA, partial cds.
 ACCESSION
 VERSION BC022604.1 GI:18490446
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1571)

AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
COMMENT Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mdedpaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

FEATURES Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRMA Plate: 43 Row: C Column: 22.
 Location/Qualifiers
 1. 1571

source
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N"
 /clone="IMAGE:4162513"
 /tissue_type="Liver, normal. 5 month old male mouse."
 /clone_id="NCI.CGAP_L19"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6"
 <1. 1493
 /codon_start=3
 /product="Similar to N-acylsphingosine amidohydrolase 2"
 /protein_id="AAH22604.1"
 /db_xref="GI:18490447"

CDS
 /translation="KLVDINGEDLGLISFAIHPIVSMNNSHNFVSDNMGVAYLFEQ
 ENKNGYLPQGFVAFSSNIGDVSPLNGHCVNTGSCNDSTCPNGSPSCMA
 SGPQDMESTHIGRIIYOKAKELVASOEVTPVLAHQVMNTDVSLNATH
 VTKCPALGYSFAAGTIDGVSGLNTGCTEGEDPMDLRDQLKPSSEIYECQPK
 PILLHSGELTIHPMPDIDVOIVTGVSLAIAIPGELTSSGRFRAIKKEFLY
 GMDKMTVIAGLSNYYTHITTYEEVOYREASRTIYSGPHLSVYIOIFRLAKAIA
 TDIVANMSSGPEPPFKNILASLIENIDRAPTRIGHFDVLOPAKREYVGEVEYIE
 VGANFRNSHENTHOTFLTEVETEDSVADWQIMNDASVETFRVHMKIGLISNATIT
 WHIPTAVPGIYRIRYFGHNRKQELIKPAVILAFEGISSPEFVVT"
 BASE COUNT 467 a 355 c 370 g 379 t
 ORIGIN

Query Match 72.9% Score 1493; DB 10; Length 1571;
Best Local Similarity 100.0% Pred. No. 0;
Matches 1493: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

557 TGAACACTGGTGGATTGATGAGAAAGACTGGGCTTATCAGCTGTTGGCATCCACC 616
 1 TGAACACTGGTGGATTGATGAGAAAGACTGGGCTTATCAGCTGTTGGCATCCACC 60
 617 CCGTAGCATGAACATAGCAACCACTTTGTGAATAGTACAATATGGCTATGGCGCTT 676
 61 CCGTAGCATGAACATAGCAACCACTTTGTGAATAGTACAATATGGCTATGGCGCTT 120
 677 ACCTTTTGGAGCAAGAAAGAAAGAGCTATCTGCTGAGACAGGAGCCGTTTGAAGCAG 736
 121 ACCTTTTGGAGCAAGAAAGAAAGAGCTATCTGCTGAGACAGGAGCCGTTTGAAGCAG 180
 737 GCTTGGCTTCATCAATCGAGAGAGCTGTACCCCAACTCTTGGCCCGCATTTGTGCTA 796
 181 GCTTGGCTTCATCAATCGAGAGAGCTGTACCCCAACTCTTGGCCCGCATTTGTGCTA 240
 797 ACACAGGGAGTCTTGTGACAGACAGAGAGCACTGTCCCAAGCGTGGGCTTACAGTGT 856
 241 ACACAGGGAGTCTTGTGACAGACAGAGAGCACTGTCCCAAGCGTGGGCTTACAGTGT 300

857 GCATGCCACGCGGACCTGGACAAAGACATTTTGAGAGCACACACATTTATGACGAGTCA 916
 301 GCATGCCACGCGGACCTGGACAAAGACATTTTGAGAGCACACACATTTATGACGAGTCA 360
 917 TCTATCAGAAAGGCAAGAGAGCTGTATGCTTGCCTGCCAGAGAGGAGACCGGCCAGTGC 976
 361 TCTATCAGAAAGGCAAGAGAGCTGTATGCTTGCCTGCCAGAGAGGAGACCGGCCAGTGC 420
 977 TTGCAGCTCAGCAGTGGGTGAACATGACAGATGTGAGCTGCAGCTCAATGGCACACACA 1036
 421 TTGCAGCTCAGCAGTGGGTGAACATGACAGATGTGAGCTGCAGCTCAATGGCACACACA 480
 1037 CAGTGAAGAGCTGTAAACCTGCCCTGCGCTACAGTTTGGCCGACGACCAATTTGATGGAG 1096
 481 CAGTGAAGAGCTGTAAACCTGCCCTGCGCTACAGTTTGGCCGACGACCAATTTGATGGAG 540
 1097 TTTGGGCGCTCAATATTATACAGAGGAAACACGAAAGGAGGATCCATTCTGGACACTCTTC 1156
 541 TTTGGGCGCTCAATATTATACAGAGGAAACACGAAAGGAGGATCCATTCTGGACACTCTTC 600
 1157 GGGACAGCTCTTGGGAAAACCATCTGAAGAGATTGTAGAGTGCAGAAACCCAAACCA 1216
 601 GGGACAGCTCTTGGGAAAACCATCTGAAGAGATTGTAGAGTGCAGAAACCCAAACCA 660
 1217 TCGTCTTACAGTGGAGAGCTGAGATACCATCTTGGCAGACCAATTTGTGATG 1276
 661 TCGTCTTACAGTGGAGAGCTGAGATACCATCTTGGCAGACCAATTTGTGATG 720
 1277 TTCAGATTGTACCGTGGGCTGCTTGGCCATGCTGCTATCCCTGGGGAATTAACACCA 1336
 721 TTCAGATTGTACCGTGGGCTGCTTGGCCATGCTGCTATCCCTGGGGAATTAACACCA 780
 1337 TGTGGGAGCAAGATTTCGTGAGCAATTAAGAAATTTGCATTTATGGAGTGAAG 1396
 781 TGTGGGAGCAAGATTTCGTGAGCAATTAAGAAATTTGCATTTATGGAGTGAAG 840
 1397 ATATGACCTTGTATTCGAGAGCTTAAGCAATGTTTATACATTTATACCATATATG 1456
 841 ATATGACCTTGTATTCGAGAGCTTAAGCAATGTTTATACATTTATACCATATATG 900
 1457 AAGAATACCAAGCTCAGCGTACGAGGAGCATCTTAATCATATGAGACACACACCGT 1516
 901 AAGAATACCAAGCTCAGCGTACGAGGAGCATCTTAATCATATGAGACACACACCGT 960
 1517 CTGCATACATCCAACTCTTCAGAGACCTTGTGAAGCAATTTGCTACGACACAGTACCA 1576
 961 CTGCATACATCCAACTCTTCAGAGACCTTGTGAAGCAATTTGCTACGACACAGTACCA 1020
 1577 ACATGAGCGTGGTCCCGGAGCTCCATCTTCAAAAATCTATACCTTCACCTATTCTCA 1636
 1021 ACATGAGCGTGGTCCCGGAGCTCCATCTTCAAAAATCTATACCTTCACCTATTCTCA 1080
 1637 ATATTGCGATGAGACCAATTTGGCAAAATTTGGGAGTCTTGCAGCCAGCAAAAC 1696
 1081 ATATTGCGATGAGACCAATTTGGCAAAATTTGGGAGTCTTGCAGCCAGCAAAAC 1140
 1697 CTGAATACAGAGTGGAGAGTGTGAAGTTATTTGTAGGCGCTAACCAAGAAAT 1756
 1141 CTGAATACAGAGTGGAGAGTGTGAAGTTATTTGTAGGCGCTAACCAAGAAAT 1200
 1757 CAGCAGAGAACCAAGCCATCAAACTTCCACGTCGTGAGAAATACGAGGACTGTGAG 1816
 1201 CAGCAGAGAACCAAGCCATCAAACTTCCACGTCGTGAGAAATACGAGGACTGTGAG 1260
 1817 CTGACTGGCAGATTAATGTTAAGAGTCCCTGTGGAGACGAGGTTTATTGGCAAAAG 1876
 1261 CTGACTGGCAGATTAATGTTAAGAGTCCCTGTGGAGACGAGGTTTATTGGCAAAAG 1320
 1877 GAATACCTGGTCTGAGCAATGCAACATATATACCTGACATATTCAGATACCTGCTACCTG 1936
 1321 GAATACCTGGTCTGAGCAATGCAACATATATACCTGACATATTCAGATACCTGCTACCTG 1380
 1937 GAATCTACAGAAATTAAGATATTTTGGACACAAATCGAAGACGAGAACTTCTGAACACCGCTG 1996

Db 321 ACTTATATACACATGATATGATATATTAACGGGTACACATATTCATTCAGTCTCTGT 380
 Oy 307 GGGTTTTCACATATACACTCTATATACCTCGCCAGCAGAGGATAGCAACCGGACCTT 366
 Db 381 GGTTTTACAGATATGATATATATGATATACAGATTAAGTTTCTATTAAGAAATTTTC 440
 Oy 367 CAGTACATAGTCTGGGATCATGAAGACATTTGATATAGTCTACACAAATCTTAACCA 426
 Db 441 GATACAAATTTGATGATATGTTCTCAACGAATTTGTAAGCTCATAAATCGGTCAACCT 500
 Oy 427 GGCAAAATCTTATATACAAAGGAATGTGCTATATGTCAGATATACCGAAGCCCTTC 486
 Db 501 GCAAAGATTTAACTCAACAGTGATATTTAGGAATTTCAAAATTAATTCGTTCCACATAT 560
 Oy 487 TCTTACTTTCGATCCACAGTACAGAGAGAGAGATTTCTTCAACACAGACAGAGA 546
 Db 561 GCATATGATATATATCCAGAGAGAAAGAAAGCATATGATGCAAAATGTTATTAAGAT 620
 Oy 547 ATGCTGCTTGAACCTGGTGAATTTGAATGAGAGAAGCTTGGTCTTATACAGTGTGT 606
 Db 621 ATGACAGTATATGATTAAGATTAAGATGATGATGATATTCATTCAGCAATTAGCTTTT 680
 Oy 607 GCATACACCCGTGATGATGATACATAGCAACCACTTTGTAATATGATACATATGGC 666
 Db 681 GGTGTTCACTTCTCATATGATATATCAAAATCTTAAATCTGTTGATTAATAAGT 740
 Oy 667 TATCGGCTTACCTTTTGGT---AGCAAGAAAGAAAGGCTATCTGCTGAGAGGA 723
 Db 741 TATGCAATTTATCTTTGGGAAAGAAATGCAATGAGCAAGCAAGCTCATTTACAGTACTGT 800
 Oy 724 CCGTTTGTAGCAGCTTTGCTTCATCAAAATCTGAGAGAGCTGACCCAACTTTGGC 783
 Db 801 CCATTCATTTGACGCTTTGGTCAAAAGTATGAAGTGTATGTTTACCAAAATCCCGTGA 860
 Oy 784 CCGCATGTGTCAACACAGGGGAGTCTGTGACACAGCAAGAGACCTGTCACAGGT 843
 Db 861 CCACATATATAGATATGATTAACCATATGTTATTAACAGTACATGTAATGACCAAGTC 920
 Oy 844 GGGCTTACATATGATGATGAGCCAGCGACCTGAGCAAGAGCTTTGAGAGACACAT 903
 Db 921 GAGAGTGTGGGCACTAGGACAGCAAGTACTGATGATGATGATGATGATGATGATGATG 980
 Oy 904 ATAGACGATCATATATACAGAGCCCAAGAGCTGTATGCTTCTCCCGAGAGGTG 963
 Db 981 ATTTGTTGATATCAATTTCAACAGGCAATTTGGAATTTATTAACAAATCAACCTTCAAGTT 1040
 Oy 964 ACCGGCCAGTGTGCTGAGCTCAGCAAGTGGGGAACATGACAGATGTGAGCTC----- 1018
 Db 1041 AGTGTAAAGATTTCAATATAGACACACATGGAACCATTTTACCAATGTTTCACTTAAGCA 1100
 Oy 1019 ---AGTCAATGCCACACACACAGTGAAGAGCTGTAAACCTGGCTGGCTACAGTTT 1074
 Db 1101 CCATATATATAGTGGGTTGAGGGCCCAACACATGATAGAGGTGCAATGGGTATATGTTT 1160
 Oy 1075 GCCGAGCAGCATATGATGAGATTTGGGCTCATATTTACACAGGAGAACTACGGAAGG 1134
 Db 1161 GCTGTGTGTACAAAGATGATGCTCAGGCTCATTTTCATTTCAAGGATGATATAGTACA 1220
 Oy 1135 GATCATTTCTGGGACACTCTTGGGACAGCTCTTGGGAAACCAATCTGAAGAGATTTGA 1194
 Db 1221 ACTATATCATTTTGAATTTTCACTGCTGATCATTTGCAAAACACACACACACAAACA 1280
 Oy 1195 GAGTGTACAGAAACCAACCAATCTGCTTCACTGAGAGAGCTGACGATACCATCTCT 1254
 Db 1281 GCTTGTCAAGCTCCAAACCAATTTTAATGTA---TGTGTGATGTTGAACCAATCCCA 1337
 Oy 1255 TGGCAGCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1314
 Db 1338 TGGGATACAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1397
 Oy 1315 ATCCCTGGGGAATTAACAACCATGCTGGGAGCAAGATTTCTGAGGCAATTTAAAAAGAA 1374

Db 1388 GTTCCAGTGAATTCACAAATATGCTGTGATGACAGATTAAGAAATACCTTATAGAAATC 1457
 Oy 1375 TTTGCATTTATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1434
 Db 1458 ATTTG---GTCAATCAATTTGAAATATGATGATGATGATGATGATGATGATGATGAT 1514
 Oy 1435 ACATATTAATTAACCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1494
 Db 1515 TCAGTATATATGCAAAATTTGAAAGATTTGCAAGTTTCAAGATTAAGAGGTGATCAAC 1574
 Oy 1495 ATCTATGACACACACACCTGCTGATGATGATGATGATGATGATGATGATGATGATG 1554
 Db 1575 GTTTTGGACCTCATCTTTAGGACCATATGCAAGAAATTTGCAAACTTTGACACATCA 1634
 Oy 1555 ATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1614
 Db 1635 ATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1694
 Oy 1615 CTAATAGCTTCACTTATTTCTATATTTGGGATGATGATGATGATGATGATGATGATG 1674
 Db 1695 CCATTTCTTTTAC---CACAGATCATTTGATGATGATGATGATGATGATGATGATG 1751
 Oy 1675 GATGCTTGGACACGCAAAACCTGATGATGATGATGATGATGATGATGATGATGATG 1734
 Db 1752 GATATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1811
 Oy 1735 GTAGCGCTTACCCCAAGATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
 Db 1812 ATTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1871
 Oy 1795 GAGAAATACAGAGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1854
 Db 1872 GATCAAT---TACATCTTAATGATGATGATGATGATGATGATGATGATGATGATG 1928
 Oy 1855 ACAGGATTTTATGACACAAAGATTAAGTATGATGATGATGATGATGATGATGATGAT 1914
 Db 1929 ACAAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1988
 Oy 1915 ATTCAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1973
 Db 1989 ATCTCCAAATCACTCAACAGGATGATGATGATGATGATGATGATGATGATGATGAT 2047

RESULT 10
 AC117816 165407 bp DNA linear HTG 16-AUG-2002
 LOCUS Mus musculus clone RP24-573C13, WORKING DRAFT SEQUENCE, 12
 DEFINITION
 ACCESSION AC117816.2 GI:22267711
 VERSION AC117816
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 165407)
 2 (bases 1 to 165407)
 REFERENCES
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
 Anderson, S., Berra, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,
 Chazaro, B., Choquel, Y., Collangelo, M., Collins, S., Collumore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S.,
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamal, A., Karatas, A., Kellis, C., Lakoque, K., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., Meidirim, J., Menees, L.,
 Mihova, T., Miengo, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,

TITLE
JOURNAL
REFERENCE
AUTHORS

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vasiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (10-Apr-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
(bases 1 to 165407)

Barren, B., Nusbam, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barron, B., Bastien, V., Bloom, T., Boguslavsky, L., Buchhalter, B.,
Cammarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diz, S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Melrim, J., Meneus, L., Mihova, T., Mjenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunhkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testfaye, S., Theodore, J., Topham, K., Travers, M., Vasiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (16-Aug-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 16, 2002 this sequence version replaced gti:20128606.
All reads were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L25854

Center clone name: 573.C13

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 160616 bases at least Q40

Consensus quality: 162867 bases at least Q30

Consensus quality: 163827 bases at least Q20

Insert size: 168000; agarose-fp

Insert size: 164307; sum-of-contrigs

Quality coverage: 7.1 in Q20 bases; agarose-fp

Quality coverage: 7.2 in Q20 bases; sum-of-contrigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 705: contrig of 705 bp in length
* 706 805: gap of 100 bp
* 806 1789: contrig of 984 bp in length
* 1790 1889: gap of 100 bp
* 1890 3396: contrig of 1507 bp in length
* 3397 3496: gap of 100 bp
* 3497 4739: contrig of 1243 bp in length

FEATURES
source
1. 165407
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RP24-573C13"
/clone_1lib="RP24-573C13"
1. 705
/note="assembly-fragment"
misc_feature
806. 1789
/note="assembly-fragment"
misc_feature
1890. 3396
/note="assembly-fragment"
misc_feature
3497. 4739
/note="assembly-fragment"
misc_feature
4840. 11073
/note="assembly-fragment"
misc_feature
11174. 18734
/note="assembly-fragment"
misc_feature
18835. 29191
/note="assembly-fragment"
misc_feature
29292. 42552
/note="assembly-fragment"
misc_feature
42653. 60018
/note="assembly-fragment"
misc_feature
60019. 60118
/note="assembly-fragment"
misc_feature
60119. 81529
/note="assembly-fragment"
misc_feature
81530. 81629
/note="assembly-fragment"
misc_feature
81630. 120500
/note="assembly-fragment"
misc_feature
120501. 120600
/note="assembly-fragment"
misc_feature
120601. 165407
/note="assembly-fragment"

BASE COUNT 46319 a 32936 c 33090 g 51958 t 1104 others
ORIGIN
Query Match 9.4% Score 192.4; DB 2; Length 165407;
Best Local Similarity 92.7% Pred No. 2e-46; 16; Indels 0; Gaps 0;
Matches 202; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
1832 TGTATACAGATCCCTCTGAGAGAGAGGTTTATTTGACACAAAGAAATGAGTCTGA 1891
1421 TTTATTTATTTGTTTCTTATTAACAGTTTATTTGACAAAGAAATGAGTCTGA 1486
1892 GCAATGCAACATTAATCTGATTCATTCAGATTAATCTGATTCATTCAGATTA 1951
1487 GCAATGCAACATTAATCTGATTCATTCAGATTAATCTGATTCATTCAGATTA 1546
1952 GATATTTTGGACACAAATGGAAGAGGAACTTGAACCCGCTGATTAATGATTTG 2011
1547 GATATTTTGGACACAAATGGAAGAGGAACTTGAACCCGCTGATTAATGATTTG 1606
2012 AAGGAATTTCTTCCTTTTGAAGTTGTCATTAATGATTTG 2049
1607 AAGGAATTTCTTCCTTTTGAAGTTGTCATTAATGATTTG 1644

RESULT 11
AC117816/c AC117816 165407 bp DNA linear HTG 16-Aug-2002

DEFINITION Mus musculus clone RP24-573C13, WORKING DRAFT SEQUENCE. 12.
 ACCESSION AC117816
 VERSION AC117816.2 GI:22267711
 KEYWORDS HTGS_P2ASEL1; HTGS_DRAFT.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 165407)
 Birren, B., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP24-573C13
 Unpublished
 2 (bases 1 to 165407)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckaghter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dehtrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferrel, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labrecque, K., Lamazares, R., Landers, T., Lehoczek, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Neill, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Plerre, N., Pollara, V., Raymond, C., Rette, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, K., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 165407)
 Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckaghter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dehtrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferrel, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Neill, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Plerre, N., Raymond, C., Rette, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (16-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 16, 2002 this sequence version replaced 91:20128606.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L25854
 Center clone name: 573_C13

----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 160616 bases at least Q40
 Consensus quality: 163827 bases at least Q30
 Insert size: 168000; agarose-fp
 Insert size: 164307; sum-of-coverage
 Quality coverage: 7.1 in Q20 bases; agarose-fp
 Quality coverage: 7.2 in Q20 bases; sum-of-coverage

 NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
 1 705: contig of 705 bp in length
 706 805: gap of 100 bp
 806 1789: contig of 984 bp in length
 1790 1889: gap of 100 bp
 1890 3396: contig of 1507 bp in length
 3397 3496: gap of 100 bp
 3497 4739: contig of 1243 bp in length
 4740 4839: gap of 100 bp
 4840 11073: contig of 6233 bp in length
 11074 11173: gap of 100 bp
 11174 18734: contig of 7561 bp in length
 18735 18834: gap of 100 bp
 18835 29191: contig of 10337 bp in length
 29192 29291: gap of 100 bp
 29292 42552: contig of 13261 bp in length
 42553 42652: gap of 100 bp
 42653 60018: contig of 17366 bp in length
 60019 60118: gap of 100 bp
 60119 81529: contig of 21411 bp in length
 81530 81629: gap of 100 bp
 81630 120500: contig of 38871 bp in length
 120501 120600: gap of 100 bp
 120601 165407: contig of 44807 bp in length.
 Location/Qualifiers
 1. 165407
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP24-573C13"
 /clone_1ib="RPCT-24 Male Mouse BAC"
 1. 705
 /note="assembly-fragment"
 806. 1789
 /note="assembly-fragment"
 1890. 3396
 /note="assembly-fragment"
 3497. 4739
 /note="assembly-fragment"
 4840. 11073
 /note="assembly-fragment"
 11174. 18734
 /note="assembly-fragment"
 18835. 29191
 /note="assembly-fragment"
 29292. 42552
 /note="assembly-fragment"
 42653. 60018
 /note="assembly-fragment"
 60119. 81529
 /note="assembly-fragment"
 81630. 120500
 /note="assembly-fragment"
 120601. 165407
 /note="assembly-fragment"

* 30962 31061: gap of 100 bp
* 31062 41639: contig of 10578 bp in length
* 41640 41739: gap of 100 bp
* 41740 77584: contig of 35845 bp in length
* 77585 77684: gap of 100 bp
* 77685 107166: contig of 29482 bp in length
* 107167 107266: gap of 100 bp
* 107267 140582: contig of 33316 bp in length
* 140583 140682: gap of 100 bp
* 140683 174914: contig of 34232 bp in length
* 174915 175014: gap of 100 bp
* 175015 184101: contig of 9087 bp in length.
Location/Qualifiers

FEATURES

source

1. 184101
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-346D12"
/clone_lib="RPC1-23 Female Mouse BMC"
1. 1186
/note="assembly-fragment"
vector_side:left"

misc_feature

/note="assembly-fragment"

vector_side:left"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

misc_feature

/note="assembly-fragment"

Db 17521 ATACTGCCAGCAGGAGGATTCAGCAACCGGACCTTTCACTACATAGTCTTGGAATCATG 17580
Oy 391 AAG 393
Db 17581 AAG 17583
RESULT 13
HUMD4C12M5 235 bp mRNA linear PRI 04-FEB-1999
LOCUS HUMD4C12M5
DEFINITION Human HepG3 partial cDNA, clone hmd4c12m5.
ACCESSION D17062.1 GI:598818
VERSION D17062.1 GI:598818
KEYWORDS gene signature.
SOURCE Homo sapiens Male cell_line:HepG2 cDNA to mRNA, clone_lib:Kiserv.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 235)
AUTHORS Matoba,R., Okubo,K., Hori,N., Fukushima,A. and Matsubara,K.
TITLE The addition of 5'-coding information to a 3'-directed cDNA library
JOURNAL Improves analysis of gene expression
MEDLINE Gene 146 (2), 199-207 (1994)
REFERENCE 94357437
AUTHORS 2 (bases 1 to 235)
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1993) Ryo Matoba, Osaka University, Institute for
Molecular and Cellular Bio: 1-3, Yamada-oka, Suita, Osaka 565,
Japan (E-mail:matoba@imh.ribo.osaka-u.ac.jp,
Tel:81-6-877-5111(ex.3314), Fax:81-6-877-1922)
Submitted (21-Jul-1993) to DDBJ by:
Ryo Matoba
Molecular Microbiology and Genetics Lab.
Research Institute of Innovative Technology for the Earth 9-2
Kizugawadai Kizu-cyo,
Soraku-gun, Kyoto
Japan, 619-02
Phone: 07747-5-2308
Fax: 07747-5-2321.
Location/Qualifiers

1. 235
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="Male"
/cell_line="HepG2"
/clone_lib="Kiserv"
BASE COUNT 64 a 46 c 57 g 68 t
ORIGIN

Query Match 7.7%; Score 158; DB 9; Length 235;
Best Local Similarity 85.8%; Pred. No. 2.4e-36;
Matches 187; Conservative 0; Mismatches 30; Indels 1; Gaps 1;
Oy 1832 TGTATACGATGCTCCCTGGAGAGAGGATTATTTGGACAAAGGATCTGGTCTGA 1891
Db 2 TGTATACGATGCTCCCTGGAGAGAGGATTATTTGGACAAAGGATCTGGTCTGA 60
Oy 1892 GCAATGCAACATATACGATGCTCCCTGGAGAGAGGATTATTTGGACAAAGGATCTGGTCTGA 1951
Db 61 GCAATGCAACATATACGATGCTCCCTGGAGAGAGGATTATTTGGACAAAGGATCTGGTCTGA 120
Oy 1952 GATATTTGGACAAATGCGAAGAGAGGATCTGAAACCCGCTGCTACTAGCATTTG 2011
Db 121 GATATTTGGACAAATGCGAAGAGAGGATCTGAAACCCGCTGCTACTAGCATTTG 180
Oy 2012 AAGGAATTTCTTCCTTTTGAAGTTGCATCTTAG 2049
Db 181 AAGGACCTCCCGGCTTTGAAGTTGTAATCTTAG 218
RESULT 14
AC012788

LOCUS AC012788 53059 bp DNA linear HTG 03-NOV-1999
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
 pieces.
 ACCESSION AC012788
 VERSION AC012788.1 GI:6223530
 KEYWORDS HTG; PHASE2.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 53059)
 AUTHORS Adams, M. and Venter, J. C.
 TITLE Direct Submission
 JOURNAL Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 COMMENT This sequence was identified as CDM:10210280 by the submitter.
 For further information on this sequence you may e-mail to
 fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1..53059
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 14592 a 11833 c 12043 g 14591 t
 ORIGIN
 Query Match 7.6%; Score 155.2; DB 2; Length 53059;
 Best Local Similarity 50.6%; Pred. No. 3.7e-35;
 Matches 640; Conservative 0; Mismatches 528; Indels 98; Gaps 7;
 QY 64 TTGATGGCTATGGCAAAATGCGCAGATACAGGGGCTCTCTCAGCAGCTGTTCAGC 123
 Db 35557 TAGATGGCTATGGCAAAATGCGCAGATACAGGGGCTCTCTCAGCAGCTGTTCAGC 35616
 QY 124 CGTCTCTTATCTTGGGGGATCCAGATGGGTCAATGCAATGGCATTTGTAGGGGTGAA 183
 Db 35617 CGTCTCTTATCTTGGGGGATCCAGATGGGTCAATGCAATGGCATTTGTAGGGGTGAA 35673
 QY 184 CTATGATATATTTCCCAAGACTGAGTTGAGGTCCTGAAAGAGACTAGAGATTAAT 243
 Db 35674 GCCGAATGATGGCTACGATGTAAGAGGAGGTGATTAAGACCTCCAGGACGTTAT 35733
 QY 244 GGCCTCTGTATCCAGAGAGCAATGTTATCTGTAGTGCATTCACAGACACTGTGGCCA 303
 Db 35734 GGCACATCTACCAATGATTAATGTGCCATCAGTCCACACGCTGCTCTCT 35793
 QY 304 GCAGGGTTTTCCTCAATATACACTATATATCTCGCCAGAGGGATTCAGACCGGAC 363
 Db 35794 GCGCGATTTCCATGATGATCTCTTATGACATCTTCATTTCTGGGCTTGTGCTCAGACG 35853
 QY 364 TTTCAGTACATAGTCTCTG-----GATCATGAG 393
 Db 35854 TTTCAGTACATAGTCTCTG-----GATCATGAG 35913
 QY 394 AGCATGATATAGCTCACACAATCTTAACACAGCAAA-----GATCATGAG 432
 Db 35914 GGAACGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 35973
 QY 433 -----ATCTTATCAACAAAGAAATGTTGTAATGTGCAGATCAACGGAAGCCCT 484
 Db 35974 ATGTCGATCTTGTCTGCTCAAAACTACTGTGCTTAATGTTAATCAATCAATGTTGCT 36033
 QY 485 CCTCTTACCTTCTGAATCCAGAGTCCAGAGAGCAAGGTATTTCTTAAACACAGCAAG 544
 Db 36034 CATCTCTACTGAGAAATCCCGCAGAGAGCTGCCAATACGAGAGCATAGCGATAGGA 36093
 QY 545 AAATGCTGCTTGAATGCTGATTTGAATGGAAGACTTGGGTTTTCACGCTGCT 604
 Db 36094 CACTGACCCAGCTGCTGATTTGTGACCTGGAAGAAACCACTCTGAGCCCTTTCAACCTG 36153

QY 605 TTGCCATCCACCCGTGAGCANTGAACATAGAACCACTTTGTGATGTGACATATAG 664
 Db 36154 ATGCGGTACATGCCACTCTCCATGAACATACCAAGACTGTGACAGCAATGTGG 36213
 QY 665 GCTATGCGGCTTACCTTTTGTAGCAAGAAAGAACAAAGGCTATCTGCTGAGAGGAC 724
 Db 36214 GTTATGCGGCTGCTGCTGGAAGAAAGATGCAATCCGAAACAGATGCCGGAAGGCA 36273
 QY 725 CGTTTGTAGCAGGCTTTCTCTTATCAATCTCGAGAGCTGTACCCCAACATTTTGGCC 784
 Db 36274 AGTTGCTGTGGCTCTTCTGCTATCCCACTTGGGAGATGTCCCCCAATTAATGGCTC 36333
 QY 785 CGCATTTGTCAACAGAGGGGAGTCTGTGCAACAGCAAGACAGCCGTCCCAAGGTG 844
 Db 36334 CCMACTGCTCATCTCCGCGCAATGATGTGATCTACTCCCTCGTTTCCACTGGCG 36393
 QY 845 GGCCTAGCATGTGCATGAGCCAGCGGACCTGAGCAAGACATGTTGAGAGCACACATTA 904
 Db 36394 AG---GGAGATGCTTGTCTCCCGGACCCGGTAAGGATATGTTGGAAGCACCCAGATCT 36450
 QY 905 TAGACGATATATATATGAGAGGCAAGAGCTG-----TATGCTCTG 949
 Db 36451 TGGGTCAACGTGTGGCGATCTCTCTGCGGACTCTCAACAGACAGCCAGAGTCCA 36510
 QY 950 CCTCCAGAGAGTGAACCGCCAGTGTCTGACGCTACACAGTGGGTGAACATGACAGATG 1009
 Db 36511 CGGACGAGGAGTCACTGGAATGTGAGGTTTCACTCCACAGTTGTGACATGCCCACT 36570
 QY 1010 TGAGGCTC-----CAGCTCAATGCGCACACACACAGTGAAGAGCTGAAC 1054
 Db 36571 ACAATGGCAGACCTACAACTCCGCTGACAGAGAGGTGACAAAGATGACAGAGGTGACG 36630
 QY 1055 CTGCGCTGGCTTACAGTTTTCGCGAGCACAATGATGAGTTTCGGGCTCAATATTA 1114
 Db 36631 CGGCAATGGCTTAATAGCTTCTCTGTGTACACAGGAGCTGAGACCTTCAGCTTCG 36690
 QY 1115 CACAGGAATACGGAAGGAGTCAATCTGGACACTCTCGGAGCAGCTGTGGAA 1174
 Db 36691 AGCAGGAAACCAACGAGCAACCCATGTGMACTTTGTGCGGAC---TTCAATCGCTG 36747
 QY 1175 AACCATGGAAGATTTAGAGTGTCAAAACCAACCAATCTGCTTCAAGTGAAG 1234
 Db 36748 CTCACAGCAGAGAGACATCAAGTGCATGAGCCCAACCATCTTCTGCGCACCGCA 36807
 QY 1235 AGCTGA 1240
 Db 36808 GGGTGA 36813
 RESULT 15
 AC008299
 LOCUS AC008299 173844 bp DNA linear INV 14-FEB-2001
 DEFINITION Drosophila melanogaster, chromosome 3R, region 99F1-99F11, BAC
 clone BACR48516, complete sequence.
 ACCESSION AC008299
 VERSION AC008299.5 GI:12830065
 KEYWORDS HTG.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 173844)
 AUTHORS Celisner, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
 Holt, R.A., Evans, C.A., Gocayne, J.D., Amaratilake, P.G., Brandon, R.C.,
 Rogers, Y., An, H., Baldwin, D., Banazon, J., Beeson, K.Y., Busam, D.A.,
 Carlson, J.W., Center, A., Chame, M., Davenport, L.B., Dietz, S.M.,
 Dodson, K., Dorsett, Y., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
 Ferreira, S., Frisze, E., Galle, R.F., Garg, N.S., George, R.A.,
 Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, R.J.,
 Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
 McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,

TITLE
Sequencing of Drosophila chromosome 3R, region 99F4-99F11
Unpublished
2 (bases 1 to 173844)
JOURNAL
GenBank
REFERENCE
Cenikler, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
Pacheco, V., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Piltman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svickas, R., Teector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
2000

FEATURES	Location/Qualifiers
source	1. 173844

BASE COUNT	48763 a	38515 c	37719 g	48847 t	
ORIGIN					
Query Match	7.6%, Score 155.2; DB 3; Length 173844;				
Best Local Similarity	50.6%; Pred. No. 4.4e-35;				
Matches 640;	Conservative 0; Mismatches 528; Indels 98; Gaps 7;				
QY	64	TTTGATGGCGTATGGCAAAAATGGCCAGAAATCAGAGGGGTCTCTCCACAGGCTGTTCACG	123		
Db	33798	TAGATGGCTATATGGCAACATCAGCAACAGTGGGTGTGGCATCCACACCCGGTCTTTGGC	33858		
QY	124	CGTGCCTTTATCTTGGGGGATCATCAGATGGCTCAATCGAATGGCATTTGTGAGCGTGGAA	183		
Db	33858	CGTGCCTTTGTGGTGGAGGATGAAAAAGGC---AACGAGTGGCCCTTGTGTAGCGCAGAT	33919		
QY	184	CTATGTATGATTTTCCCAACGACTGAGTTGGAGTCTCTGAAGAGACTAGAGATTAATAT	243		
Db	33915	GCCCGAATGATGGGCTACGGATTTGAAGGGAGGTGATTAACAGCACTCCAGCAGCTTAT	33979		
QY	244	GGCTCTCGATATGCAAGAGACAAATGTTATCCTGATGGCATTCACACACTCTGGGCCA	303		
Db	33975	GGCAACATCTACCAACAAATGATTAATGTGGCCATCAGTGGCACCCACGACGAGGTCTCT	34033		
QY	304	GCAAGGTTTTTCCAAATATACACTCTATATATCTGCCACGAGGAGTTGACCAACCGGACC	363		
Db	34035	GGCGGATTTCCGATGAGCATCTGCTCTATGACATCTCCATCTCGGCTTTGTGCTCCAGACG	34099		
QY	364	TTTTCAGTACATAGTCTCTGG-----GATCATGAAG	393		
Db	34095	TTTGAAGTATGGCTCAGGAGACTATCTGTGGGTTTTCACATTCACACTGAATTTGTAC	34155		
QY	394	AGCATTTATATAGTCCACACAATCTTAACACAGGCAAA-----	432		
Db	34155	GGAAACGAATTAATTAATTAATTAATAGTGCATCAAAAAAGGCGACGACAACTGTGG	34219		

OY	433	-----ATCCTTAATCAAAAGGAAATGTTCCTAATGTGCAGATCAACCCGAAAGCCCT	484
Db	34215	ATGTCGCATCTCTCTCTCCAAAATCTAGTCTGCTAAATGTAACTAATCAATCGTTCGCCCT	34274
OY	485	CCTCTTACCTCTTGAAATCAACAGTACAGAGAGACAAAGTAAATCTTCAACACAGACAAG	544
Db	34275	CATCTCACTTGAGGAATCCGGCGAGGAGGCTGCCCAATTCAGAGCAAGATACGGATTAAAG	34336
OY	545	AAATGCTGTCCTTGAACCTGGTGGATTTGAAATGAGAGAAAGCTTGGCTTTATACGTGGT	604
Db	34335	CACGAGCCCAAGCTGCGATTTGTGTGACCTGGAAAAACAACCTCCGCGGCGCTTCAACCTGGT	34394
OY	605	TTGCGATTCACCCCGTGAAGATGAACAATAGCAACCACTTGTGAATAGACAAATATG	664
Db	34395	ATGGGGTGCAATGCCACCTCCATGAAACAATACCAAGACACTGGTGACAGAGCAAAATGTGG	34455
OY	665	GCTATGCGGCTTACCTTTTGTAGCAAGAAAAAGAAACAAAGGCTATGTGCTGACAGGAC	724
Db	34455	GTTATGCGCGCTGCTCTCTGGAAAAAGATACATCCGAACAAAGATGCGCGGAAAGGCA	34514
OY	725	CGTTGTAGAGAGGCTTGTGCTTCAATCAATCCGAGACGTGACACCACATCTTGGCC	784
Db	34515	AGTTTCGTGGGCTTCTGTGCTATCCACTTGGCGATGTGTGCTCCCAATATATATGGGTC	34574
OY	785	CGCATTTGTTCACACACAGGGGAGTCTTGTGACAACAGCAAGACACCTGTCCCAAGGTC	844
Db	34575	CCAAGTGTGTGATCTCGCGCAATGAGATGTATCTAGTACCTCCGTTGGCCACATGGGG	34634
OY	845	GGCTACATGTGATGGCCAGCGGACCTGTGACAACATGTTTGGAGACACACATTA	904
Db	34635	AG---GGAGATTTGTTTGGCTCCGACCCGGTAAGGATATGTTGAAAAGCCCAATCT	34691
OY	905	TAGACGGAATCATATATCAAGAGCCCAAGAGCTG-----TATGCTCTG	949
Db	34692	TGGTTCAAAGCTGTGGCGATGTGCTGTGGGACTGTCCAAACGAAACAGAGCAAGATCCA	34751
OY	950	CTTCCAGAGGTGACCGGCCAGTGTCTACAGTCCACAGTGGGTGAACATGACAGATG	1009
Db	34752	CGGACGCGGAGGTCACTGGAGATGTGATTCATCCACCGATTTGTGGACATCCCAACT	34811
OY	1010	TGACCTGCT-----CAGGTCAATGCACACACACAGTGAAGAGCTGTAAAC	1054
Db	34812	ACAATGGCAGACCTACCAATCCGTGTAGCAGGAAGTGCACAAAGATCAGGGATGTGCAC	34871
OY	1055	CTGCCCTGGGCTACAGTTTGTCCGACAGCACAATGTATGAGATTTGGGGCTCAATATTA	1114
Db	34872	CGGCAATGGGCTAATAGCTTGTGCTGTGTACCAACCGATGGAACCTGAGGCTTACGCTTG	34931
OY	1115	CACAGGGAACATAGAGGAGGGAATCCATCTTGGGACACTGTCTGGGACACAGCTTGGGA	1174
Db	34932	AGCAGGGAACACACAGGACACACCCCATGTGTGAATCTTTGTGCCGAC---TTCATGCTG	34988
OY	1175	AACCATCTGAAGAGATTGTAGAGTGTCAAGAAACCAACCAATCTGCTTCACTGAGAG	1234
Db	34989	CTCCACGACGAGGAGACATCAAGTGCATAGAGCCCAAGCCCATTTCTTGGCCACCGGCA	35048
OY	1235	AGCTGA 1240	
Db	35049	GGGTGA 35054	

Search completed: July 5, 2003, 11:30:33
Job time : 5399 secs


```

Oy 362 TIDVSGNTINOGTTGEGPMDTLRDOLGPPSEIVEICQKRPILLHSGELTIPHPWQ 420
Db 358 STEDPGPLGLEG--NNPFLSALGGLLTGPPOCLVQCAEKTLLADTGNKK-PPWPT 413
Oy 421 PDIVVOJIVTVGSLAIAIPGELTITMGSRRPREAIKEEFALYCMKDMTVIAGLSNVTYH 480
Db 414 PTVLPIDMFRIQOGLLELGPAPAEFTYMACGRIRRAVOAASEAMGIRH--VYFNCTANAYAS 471
Oy 481 YTTYEEXYQOORYEASTTYGPHLTLSAYTQLEFRDLAKA-----IATDTVA-----N 526
Db 472 YTTTREETAAQREYEGSGSTLYGPMTOQAAYOQLFVDMAVALRERLPVETSAIAPDLSCOMN 531
Oy 527 MSSGGEPPPEFKULIASLPINADBRAPICGNFGEDVLOPAKPEYRVEGEVEYETIPGANKNS 566
Db 532 FOTG-----VADDPYIGKSFSDVLOQPPRESYVIGCKVVAFTGPHKN- 575
Oy 587 AENQHOHFLVEXY-EDSVADMOQIMYNDAASMETRPMYHKGILGLSNATYVWHIIPDTAYP 645
Db 576 -DLRKEKTEFLEVYNIKGDKQOTPVYATVDMDTQYRMERBVGISASKATISWSIIPOTEP 634
Oy 646 GYIRIRYFGHNKQOELLKPAVILAEGISPFEVYTT 682
Db 635 GHYITRHYGNMAKNFTQK---ISEIGGSRSEVLT 668

```

RESULT 2
 US-09-328-501-1
 : Sequence 1, Application US/09328501A
 : Patent No. 6258581
 : GENERAL INFORMATION:
 : APPLICANT: OKINO, No. 6258581omn
 : APPLICANT: ITO, Makoto
 : TITLE OF INVENTION: Ceramidase Gene
 : FILE REFERENCE: 1w22-0377p
 : CURRENT APPLICATION NUMBER: US/09/328-501A
 : CURRENT FILING DATE: 1999-06-09
 : EARLIER APPLICATION NUMBER: 10-234769 JAPAN
 : EARLIER FILING DATE: 1998-08-20
 : NUMBER OF SEQ ID NOS: 18
 : SOFTWARE: Patentin Ver. 2.1
 : SEQ ID NO 1
 : LENGTH: 645
 : TYPE: PRT
 : ORGANISM: Pseudomonas aeruginosa
 : FEATURE:
 : OTHER INFORMATION: any n or xaa = Unknown
 US-09-328-501-1

[illegible]

```

QY      304 RIYYQAKAEYVYASASEVYGPVLAAVHNNMIDVSVOLNAT - -HYVTKCKPALDGSPAA 361
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      273 LRQFAKAYELAQADQBEVUGELDSFRFYVDTRLP LRP EFTDQGPQLCTAAIGTSLAAG 332
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      362 -TIDVSGSINITQGTGGDPFMDLRDOLLKPSBEIVECQKPKPILLHSGELTIHPWQ 420
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      333 STEDGCGPLGEBG - -NNPFLSALUGELLTGVPQPELVQCAEKTLLADIGNKK - PYPWT 388
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      421 PDIVDQIVTVGSLAIAIPGELITMSGRRRRERAAIKKEFALYGMKKMTVYIAGLSNYTH 480
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      389 PLYLPLOMRIGOLELTLGAPAEFTYMAQVYRIRRAQAASEAAGIRH - -VYFNAGNAYAS 446
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      481 YTYTEEYQOAREASTYGPHTLSAYITQIFRLDIA - - - - -IATDTVA - - - - -N 526
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      447 YVTTREYAAQEEGGSTLGPWTQAATQQLFVDMAYALRRLRELVETSATLRDLSCCQM 506
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      527 MSSGPBPFFKMLIASLRPNIDRAPRIGKHEGVDLPQAPREYRVEGEVVEVLYFGANPKNS 586
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      507 FQGVG - - - - - - - - - - -VADDPYIGKSGFDVLDQQRRESYRIGDKVYTAFAVGHPNK - 550
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      587 AENQTHQTELYEYK - EDSVADWQIMYNDASHEFRFYNNKSLGLSNATYIMHPDTAYP 645
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      551 -DLRTEKTELEVANNIGKDGKQTPVAVDNDMDQOYRWERGVISASKATISWSLIPQTEP 609
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      646 GIVIRYFGNRRQKELLKPAVLAREGISSPPEVYTT 682
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      610 GHYIRHYGAKKFWTQK - - - - -ISLIGSGSTRSEFVLT 643
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 3
US-09-468-578-4
Sequence 4, Application US/09468578
Patent No. 6399329
GENERAL INFORMATION:
APPLICANT: Wang, Huaming
APPLICANT: Bodie, Elizabeth A.
TITLE OF INVENTION: Phenol Oxidizing Enzymes
FILE REFERENCE: GC561-3
CURRENT APPLICATION NUMBER: US/09/468,578
CURRENT FILING DATE: 1999-12-21
PRIORITY APPLICATION NUMBER: US 09/220,871
PRIORITY FILING DATE: 1998-12-23
PRIORITY APPLICATION NUMBER: US 09/338,723
PRIORITY FILING DATE: 1999-06-23
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 627
TYPE: PRT
ORGANISM: Bipolaris spicifera
US-09-468-578-4

```

Query Match	3.1%;	Score 11.5;	DB 4;	Length 627;
Best Local Similarity	20.8%;	Pred. No. 0.02;		
Matches	129;	Conservative	76;	Mismatches 230; Indels 105; Gaps 35
QY	145	IFIKKGVANVAQINRSPESYLLN--PQSERARYSS--NTDKEMLYKLVLDNGEDLGLT--	199	
Db	13	VSIAKQ--IYGVALSERPAFDNTPDEKKAALASIVEDDPADVMMMLDWMOSPEXPLIFR	71	
QY	200	SWFAIHPVSMNN--SNHFNVDNMGYAUL--FEQEKMKGYLPGCGPFPVAGFASNNLADV	255	
Db	72	QPLFIIPRAKEKNKLNPVATNKEIWTYELVIRPFQGVYPSLRPAR---LVGIDGISTPG--	126	
QY	256	SPNLTGP-----HCVNTGESCDNDKSTCPNGSGSMCASGPGQDMFESTHIIIGRIITYQ	308	
Db	127	-PRTIIVRGTEAVVYRFINOG---DRESSIIHLHGSPSRAPFGDMADM-----IMKG	173	
QY	309	KAKELIVASAGEVATGVPYLAHQVNMNTDVSYQLNATHVTKCKPALGVSFAAGTIDGVSQ	368	
Db	174	EYKRYIYPPNNQ-----AARFLMTHD-----HAMHT-----AENATFGQAG	209	
QY	369	LNITQGTTEGDFEMDTLRDQLLKPS---EIVVECCQPKPILLHSGELTTPHPMQDPIV	424	

Db 210 AYLT-----DPAEDA-----LGLPSGKDYI-----PLVLSKYYNADGTLKTSVG 252

QY 425 DVQVTVGSLAIAIPELITMSGRFR-----EAIKERRALYMK-DMYV-----VIA 472

Db 253 EDSWSMGDIIHNGOQPMPLINVEPRKYLRFELNAAVSRNALYFVKODMNTRLPFOVIA 312

QY 473 GLSNVYHYITTYEY--OAOXYEAASTI--YGPHTLSAYITQFRDLAKA-----518

Db 313 SDAGLTHPVQTSIMYVAAAEFRITVDFAPYAGOTID-----LRNKAANGITDDDYA 367

QY 519 -----IATDTVANMSSGPEPFKNIASLIPIADRAPIGKHGVLPAKPEYR 569

Db 368 NTDKVMRFHVSQGVVNSVPEQ-----LSQIQPPADKTDIDHHR--FHRTNGEMR 418

QY 570 VGEY-----VEYIVGANPK-----NSAENOTHO-----TFLVEKTEDSVADMOI 610

Db 419 INGIGFADVENRYLAKVPGTVELMELENSGSGSHPIHVLVDFRVVARYGDEGTGVM 478

QY 611 MY-----NDASWETRFYMHGILGLSNATYWHIPDTAVPGIY-----648

Db 479 PYEAGLADYVWLGR---HETVL-----VEAHY--APWDGVYFHCNLIHEDQDMAA 527

QY 649 ----RIRYFGHNKQELLP 664

Db 528 FDVTKLQNFQYNETTDFHDP 547

RESULT 4

US-08-219-262B-9

Sequence 9, Application US/08219262B

Patent No. 5788970

GENERAL INFORMATION:

APPLICANT: VAKHARIA, VIKRAM

APPLICANT: SNYDER, DAVID B

APPLICANT: MENDEL-WHERSAT, STEPHANIE A

TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS

TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED

TITLE OF INVENTION: THERON

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT

STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/219,262B

FILING DATE: 29-MAR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 2747-047-27

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 1012 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Infectious bursal disease virus

STRAIN: 002-73

US-08-219-262B-9

Query Match 3.0%; Score 108; DB 1; Length 1012;

Best Local Similarity 20.9%; Pred. No. 0.11;

Matches 104; Conservative 72; Mismatches 192; Indels 130; Gaps 26;

QY 231 EKNNGVLPQGPFPVAFASS-----NLGDSVPII-LGPHCVMTGESCDNDK-----275

Db 153 DKIGNVLVGEVTVLSLPTSIDLGYRLGDPIDPAIGLDPKVVATCSDRRVYTTIRAD 212

QY 276 -----STCPNGSPSCMAAGPGQDMFESTHILIGRIIYOKAKELVYASOEYTGVLAAHQ 330

Db 213 DYQFSQYQPGCVITLIFSA-NIDAINSLVSGELVPEQTS-----VQGLVL-----257

QY 331 VWNKTDVSVOLNATHYKTCYPALGYSFAGSTIDVSGNLITGCTGEGDPWTLRQOL 390

Db 258 --NATITLVGFDGT-TVTTRAVAAAGNGITAGT-DNLMPFNLVPTSETIQPVTSIKLEIV 313

QY 391 -----GKPSREIVCQPKPILHSGELTI-----PHMQP-DIVDVQIVTVGSLAIA 437

Db 314 TSKSGGQAGDM-----SWLASNLAVITHGNTPGALRPVTLVAYEVANGSVYTV 365

QY 438 AIPGELITMSGRFRREAIKKEFALY--GMKDMTVI-----AGLSNVY-THYITTYEY 488

Db 366 AGVSNFELIPNELAKNLVTEYGRDPGAMVTKLILSERDRLGKTYMPFREYTDREY 425

QY 489 OAOXYEAASTYIGHTISATITQFRDLAKAIAITDTVANMSSGPEPFKNIASLIPIA 548

Db 426 ---FMEVADLNSPLKTAGAG-FKDIIRAIRIAYP-----VSTLFP---464

QY 549 DRAPIGHFGDVLDPAPREYVGEVEYIF-----VGANPKSAENOTHOFLTEYKEDS 604

Db 465 PAFLAIAIGGV-----DYLGDPAQASSTAASGKARAAAGRIROLTLADKGYEV 519

QY 605 VAD-WQIMYNDASMETRFYMHGILGLSNATYWHIPDTAVPGIYRIYFGHN-----RK 658

Db 520 VANLFYQPNVY-----DGIL-----ASPVLR---GAHNDCVIRE 554

QY 659 QELKPAVITLAFEGISSP 676

Db 555 GATLEPVVITTVEDAMTP 572

RESULT 5

US-09-031-655-9

Sequence 9, Application US/09031655

Patent No. 6017759

GENERAL INFORMATION:

APPLICANT: VAKHARIA, VIKRAM

APPLICANT: SNYDER, DAVID B

APPLICANT: MENDEL-WHERSAT, STEPHANIE A

TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS

TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED

TITLE OF INVENTION: THERON

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT

STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/031,655

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/219,262
FILING DATE: 29-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORGANISM SOURCE:
ORGANISM: Infectious bursal disease virus
STRAIN: 002-73
US-09-031-655-9

Query Match 3.0%; Score 108; DB 3; Length 1012;
Best Local Similarity 20.9%; Pred No. 0.11;

Matches 104; Conservative 72; Mismatches 192; Indels 130; Gaps 26;

QY 231 EKNKGYPGQGPFFVAGFASS-----NLGDVSPNI-LGPHCVNTGSCDNDK----- 275
DB 153 DKIGNLVGEGVTLSPTSYDGYRLGDPPIALGDPKMWATCQSSRPRTYTTAAD 212
QY 276 -----STCPNGSPSCMASGPGQDMFESHTIIGRIYQAKELVYASASOEVTGPVLAHQ 330
DB 213 DYQFSQYQPGGVITTLFSA-NIDAINSLVSGELVEFQTS-----VOGLVY----- 257
QY 331 WYNNMDSVQNLNATHVKTCKPDLGYSFAAGTIDVSGLNTGTEGDPFMDTLBDOLL 390
DB 258 --NATLYLVGFGT-TVTRAVAAAGNLTAGT-DNLMPNLVPTSETIQPTSTILETV 313
QY 391 -----GKPESEIVECQPKRPILLHSGELTI-----PHWPQ-DIVDQVITYVSLAIA 437
DB 314 TSKSGQAGDDQ-----SWLASGNLAVTTHGNYPCALRPVTLVAVERVATGSGVTV 365
QY 438 ALPGELTTMGRFRRAIKKEFALY--GKMDTVI-----AGLSNV-THKITTYEEY 488
DB 366 AGVSNFELLPNBELAKNLVTEYGRDPMNMTKLLSRDRIGITVMPTRYDREY 425
QY 489 QAQRREASTYCPHTLAVIQLFRDLAKAIAITDVANSSGPEPEFKNLASLIPNIA 548
DB 426 -----FMEVADLNSPLKIAAGF-FKDIHAIKRIAVP-----VSTLFP----- 464
QY 549 DRAPIGKHEGVDLQAPKPEYRVEVEVIE---VGANPKNSAENQHTQFTLVEKEDS 604
DB 465 PAAPLHAHAGEGV-----DYLLGDEQAASGTAARAASGRIRQLTLAADRGYEV 519
QY 605 VAD-WQIMVNDASWETREYMHKGIGLSNATYIMHPDPAVPGIYIRIRFGH-----RK 658
DB 520 VANLEQVQPNPVY-----DGIIL-----ASPGVLR--GAHNLDCVLR 554
QY 659 QELKPAVLALEGISSP 676
DB 555 GATLEPPVITTYEDAMTP 572

RESULT 6
US-08-570-311-14

Sequence 14, Application US/08570311

Patient No. 5824791

GENERAL INFORMATION:

APPLICANT: Prologis-Kee-Fox, Ann

APPLICANT: Tumwasorn, Somying

APPLICANT: Lepline, Guyelaine

APPLICANT: Hen, Nalinang

APPLICANT: Lantiz, Marilyn

APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSER: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UP15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2628 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-14

Query Match 3.0%; Score 106.5; DB 2; Length 2628;
Best Local Similarity 18.8%; Pred No. 0.94;

Matches 116; Conservative 82; Mismatches 236; Indels 183; Gaps 31;

QY 101 PAGE--FOYTYTILASEGFSNRTFOYIVSGIMK-----SIDIAHTNLKPKGIFT--- 147
DB 105 PAGLYDPFEYKVPVNDASFSPTNF--VLDGTAADIPAGTYDVIYINPNPGLIYVGG 162
QY 148 -NKGNAVQINRSPSYLINPOSERARISNTDKEMLVKLVLDNGELGLISMFALHP 206
DB 163 VSKGNDYVEAK--TYHFTVQROGP-----GDAASYV----- 193
QY 207 VSNMNSNHFVNSDNGYAAVLEFQEKNKGYLPGQGFVAGFASSNLGDVSPNIGPHCVN 266
DB 194 VTGEGNEFAPVONLMSV-----SCQTVTLTWQAPASDRITYL----- 233
QY 267 TGESCDNDKSTCPNGSPSCMASGPGQDMFESHTIIGRIYQAKELVYASASOEVTGPVL 326
DB 234 -NESFDT--QTLPGNG-TMIDADGCGHNLSTIN-----YVTAHTGSG-AM 276
QY 327 AAHQVNMDSVQNLNATHVKTCKPA-----LGSPAGTIDGVS---GLNITQCTTE 377
DB 277 FSKSWTASGAKIDISPDNVLVTPKTVYDENGKLSY-WVSSQVPTNEHYGVFLSTGNE 335
QY 378 GDPFMDTLBDOLLGKSEIVECQPKPI-LHSGELTIPHWPQDIVDQVITYVSLAIA 436
DB 336 AANFTIKLEETLGS-----DKPAPNLVKSQVGLPAPYQERTID-----L 377


```
REFERENCE/DOCKET NUMBER: B0801/7023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 878 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-732-429-2

Query Match      2.9%; Score 104; DB 4; Length 878;
Best Local Similarity 18.6%; Pred. No. 0.23;
Matches 130; Conservative 94; Mismatches 255; Indels 220; Gaps 32;

OY 47 ADPOGSMFAVSVELCMISQRLREY-----LKR-----LESYXG 82
DB 41 APPERRPLGRVNEEDCTGRORTAIFLPIPKVGTGDIYIVKRLPLRPHNPDPFLGLKMD 100
OY 83 SLYRR--DNVILSAIHTHSGPAGFEQYLYILASEGSENRFTFOYIVSGIMKSIDIAHTNL 140
DB 101 STYRKFSKVTILNTVGHHRP-----PHQASVSGIOAEL-LTFPNS 141
OY 141 KPGKIFINKGVANVOIKRSPSYLLN---POSERARY-----SNTDKEMLYLKV 189
DB 142 SPG-----LRRORRWVPIIPISCPENKGFPRKLVQIKSNKKEGVFYSI 188
OY 190 DLNSEDGLISMFAI-----HPVSNNSNHFVNSDNMGYA 224
DB 189 TGCGADTPPVGVFIETRETGLKTEPLDRERIAITTLFSAVSSNGN---AVEDPMEL 245
OY 225 AYLFEQENK-----GYLPQGPVAVGAFASNLGDVSPNLLGPHCVNTGES 270
DB 246 ITVDQNMNKBEPFOEVFKGSVMGALPG-----TSVMEV-----TATD 284
OY 271 CDNKSCTPCNGSPSCMASG---PGQDMF---ESTHILGRIITYOKAKELASASQEVYTG 323
DB 285 ADDDVNTY-NAIAVTIISDPELPDKMFTINRGTGISVYTTGLDRESPPYTLVVQA 343
OY 324 PVLAHQVNMNTDVSVOLNATHVTKCPALGYSFAACTIDG-----VSGLNTQ 373
DB 344 ADLOG-EGLSITATAV---ITVTDNDNRP---FNPTTYKGVPEANNAVITTLAKYTD 396
OY 374 GTTGDDPFWTLRDQLGKPSSEIVECOKP-----KPIILHSGELTIPHPWQDI 423
DB 397 ADAPNTPAMEAVY-TILNDGCGQFVYTNPNVNDGILTKAGLDPEAKQVILHVAVTNV 455
OY 424 VDVOIVYVGSIAIAIPBELITMGGRR-----REAIKKEFALYGMKMTVYIAGLSVY 478
DB 456 VPFEVSLTSTAIVTV-DVLVDNMGPIFVPEKREVSSEDFGV-GQETSYTQOEPDTF 512
OY 479 THYITVEYOQOREASTIYGPHT--LSAVIOLFDR-----LAKAIAITDVYAN 526
DB 513 MEQKITIYIMWDRNMLE---INPDGAISTRAEIDREDFEHVKNSTYITALIINTDNSP 569
OY 527 MSSGPEPPFFKNIASLIPNIADRAPIGKHEGVDVLOPAKPEYRGEVEVEIVGANKP-- 584
DB 570 VATG-----TGTLLILSDVNDNAPI-----PEPRT-----IFECERNRPQ 606
OY 585 -----NSAENQTHQFLTYEKEDSVADQIMYNDASMET 619
DB 607 VINIHADLPNTPSPFTALTHGRVPMWTIOYNDPTQES 645

RESULT 9
US-09-798-267-2
Sequence 2, Application US/09798267
Patent No. 6406870
GENERAL INFORMATION:
APPLICANT: Brenner, Michael
APPLICANT: Cepdek, Karlyn
TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
```

```
TITLE OF INVENTION: Interactions with T Lymphocytes
FILE REFERENCE: L0560/7008ERP
CURRENT APPLICATION NUMBER: US/09/798,267
CURRENT FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: US 08/237,919
PRIOR FILING DATE: 1994-05-03
PRIOR APPLICATION NUMBER: PCT/US 95/05518
PRIOR FILING DATE: 1995-05-03
PRIOR APPLICATION NUMBER: US 08/732,429
PRIOR FILING DATE: 1996-11-01
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 878
TYPE: prt
ORGANISM: Homo sapiens
US-09-798-267-2

Query Match      2.9%; Score 104; DB 4; Length 878;
Best Local Similarity 18.6%; Pred. No. 0.23;
Matches 130; Conservative 94; Mismatches 255; Indels 220; Gaps 32;

OY 47 ADPOGSMFAVSVELCMISQRLREY-----LKR-----LESYXG 82
DB 41 APPERRPLGRVNEEDCTGRORTAIFLPIPKVGTGDIYIVKRLPLRPHNPDPFLGLKMD 100
OY 83 SLYRR--DNVILSAIHTHSGPAGFEQYLYILASEGSENRFTFOYIVSGIMKSIDIAHTNL 140
DB 101 STYRKFSKVTILNTVGHHRP-----PHQASVSGIOAEL-LTFPNS 141
OY 141 KPGKIFINKGVANVOIKRSPSYLLN---POSERARY-----SNTDKEMLYLKV 189
DB 142 SPG-----LRRORRWVPIIPISCPENKGFPRKLVQIKSNKKEGVFYSI 188
OY 190 DLNSEDGLISMFAI-----HPVSNNSNHFVNSDNMGYA 224
DB 189 TGCGADTPPVGVFIETRETGLKTEPLDRERIAITTLFSAVSSNGN---AVEDPMEL 245
OY 225 AYLFEQENK-----GYLPQGPVAVGAFASNLGDVSPNLLGPHCVNTGES 270
DB 246 ITVDQNMNKBEPFOEVFKGSVMGALPG-----TSVMEV-----TATD 284
OY 271 CDNKSCTPCNGSPSCMASG---PGQDMF---ESTHILGRIITYOKAKELASASQEVYTG 323
DB 285 ADDDVNTY-NAIAVTIISDPELPDKMFTINRGTGISVYTTGLDRESPPYTLVVQA 343
OY 324 PVLAHQVNMNTDVSVOLNATHVTKCPALGYSFAACTIDG-----VSGLNTQ 373
DB 344 ADLOG-EGLSITATAV---ITVTDNDNRP---FNPTTYKGVPEANNAVITTLAKYTD 396
OY 374 GTTGDDPFWTLRDQLGKPSSEIVECOKP-----KPIILHSGELTIPHPWQDI 423
DB 397 ADAPNTPAMEAVY-TILNDGCGQFVYTNPNVNDGILTKAGLDPEAKQVILHVAVTNV 455
OY 424 VDVOIVYVGSIAIAIPBELITMGGRR-----REAIKKEFALYGMKMTVYIAGLSVY 478
DB 456 VPFEVSLTSTAIVTV-DVLVDNMGPIFVPEKREVSSEDFGV-GQETSYTQOEPDTF 512
OY 479 THYITVEYOQOREASTIYGPHT--LSAVIOLFDR-----LAKAIAITDVYAN 526
DB 513 MEQKITIYIMWDRNMLE---INPDGAISTRAEIDREDFEHVKNSTYITALIINTDNSP 569
OY 527 MSSGPEPPFFKNIASLIPNIADRAPIGKHEGVDVLOPAKPEYRGEVEVEIVGANKP-- 584
DB 570 VATG-----TGTLLILSDVNDNAPI-----PEPRT-----IFECERNRPQ 606
OY 585 -----NSAENQTHQFLTYEKEDSVADQIMYNDASMET 619
DB 607 VINIHADLPNTPSPFTALTHGRVPMWTIOYNDPTQES 645

RESULT 10
US-09-798-267-3
```

Sequence 3, Application US/09798267
Patent No. 6406870
GENERAL INFORMATION:
APPLICANT: Brenner, Michael
APPLICANT: Ceprek, Karyn
TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
FILE REFERENCE: L0560/7008ERP
CURRENT FILING DATE: 2001-03-02
PRIORITY FILING DATE: 1994-05-03
PRIORITY FILING DATE: 1995-05-03
PRIORITY FILING DATE: 1996-11-01
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 878
TYPE: PRT
ORGANISM: Homo sapiens
US-09-798-267-3

Query Match 2.9%; Score 104; DB 4; Length 878;
Best Local Similarity 18.6%; Pred. No. 0.23;
Matches 130; Conservative 94; Mismatches 255; Indels 220; Gaps 32;

47 ADPDGSRMAFVSELCMISQRLLEY-----LKR-----LESKYG 82
41 APEKRPRLGRVNEEDCTGKQRTAIFLTPKVGTDGVITVKRPLRHPNTPDPLGLRWD 100
83 SLVRR--DNVILSAIHTSGPAGFFOYTYLLASEGFSNRTFOYVSGIKSIDIANLNL 140
101 STYRKSTXTKTLTVGHHRRP-----PHQASVSGIOAEL-LTRPNS 141
141 KPKKIFLNKGNVANVOJNRSPSSYLLN-----POSERARY-----SSNTDKEMLYLKV 189
142 SPG-----LRQKRDVWIPRISCPENKGPFRKNLVQIISNNDKEKVFYSI 188
190 DLNGEDLGLISWPAI-----HPVSMNSNHFVSDNNGYA 224
189 TGQCADTPPVGVFLIERETGWLKTEPLDRERIATYTLFSAVSSNNGN--AVEDPMEIL 245
225 AYLEDEKKNK-----GYLPGGPVAGFASNLGDVSPNIIIGPHCVNNGES 270
246 ITVTDQDNKPEFTQEVFKGVSMEGALPG-----TSVMEV-----TATD 284
271 CDNDKSTCPNGGSPMCMAAG-----PGQDMF-----ESTHIIIGRIIYQAKELIYASASQEVYTG 323
285 ADDVNTY-NAALAYTILSDPELPKMNFTINRNTGCVISVTTGLDRESFPYTYLVQA 343
324 PYLAHOMVMTDVSVQNLNTHVTKCPALGYSFAAGTIDG-----VSGLNITQ 373
344 ADLQG-EGSLTATAV---ITVTDNDNPPF---FNEPTYKGOVPEANEAVVITTLKATD 396
374 GTTEGDFMDTLDDQLGKPSSEELVEQKP-----KPIILHSGELTIPHPMPDI 423
397 ADAPNTPAMEAVY-TIINDGGOFVVTTFNVNNDGILKTAKGLDTEKKQOYILHVAVTNV 455
424 VDVOIYTVGSLAIAIPGELTJMSGRF-----REAIKKEFALYGMKDMVTVIAGLSNVY 478
456 VPPEVSLTSTAIVTV-DVLDVNEGPIFVPEKREVESEDFGV--GQELISYTAQEDDTF 512
479 THITTYEYQAOYEASTTYGPHN--LSAVTQLFRD-----LAKATATDTVAN 526
513 MEOKIYIRWRDRNMLE---INPDGAISTRALREDEHVKNSTYTALITATDNGSP 569
527 MSSGPEPPFKNLIASIPNIADRAPIGKHFQVLPAPKPEYRAGEVEVFIYGANKP-- 584
570 VARG-----TGILLILSDVNDNAPF-----PEPRP-----IIFCERNKPKQ 606
585 ---NSAENQTHQTLFVEKYEKDSVADMOIMYNDASWET 619

Db 607 VINIHADLPNTPSPFTAEILHGRVPMWITIOYNDPTGES 645

RESULT 11
PCT-US95-05518-2
Sequence 2, Application PC/TUS9505518
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
FILE REFERENCE: E-cadherin interactions with T Lymphocytes
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05518
FILING DATE: herewith
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/237,919
FILING DATE: 3 May 1994
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 878 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-05518-2

Query Match 2.9%; Score 104; DB 5; Length 878;
Best Local Similarity 18.6%; Pred. No. 0.23;
Matches 130; Conservative 94; Mismatches 255; Indels 220; Gaps 32;

47 ADPDGSRMAFVSELCMISQRLLEY-----LKR-----LESKYG 82
41 APEKRPRLGRVNEEDCTGKQRTAIFLTPKVGTDGVITVKRPLRHPNTPDPLGLRWD 100
83 SLVRR--DNVILSAIHTSGPAGFFOYTYLLASEGFSNRTFOYVSGIKSIDIANLNL 140
101 STYRKSTXTKTLTVGHHRRP-----PHQASVSGIOAEL-LTRPNS 141
141 KPKKIFLNKGNVANVOJNRSPSSYLLN-----POSERARY-----SSNTDKEMLYLKV 189
142 SPG-----LRQKRDVWIPRISCPENKGPFRKNLVQIISNNDKEKVFYSI 188
190 DLNGEDLGLISWPAI-----HPVSMNSNHFVSDNNGYA 224
189 TGQCADTPPVGVFLIERETGWLKTEPLDRERIATYTLFSAVSSNNGN--AVEDPMEIL 245
225 AYLEDEKKNK-----GYLPGGPVAGFASNLGDVSPNIIIGPHCVNNGES 270
246 ITVTDQDNKPEFTQEVFKGVSMEGALPG-----TSVMEV-----TATD 284
271 CDNDKSTCPNGGSPMCMAAG-----PGQDMF-----ESTHIIIGRIIYQAKELIYASASQEVYTG 323
285 ADDVNTY-NAALAYTILSDPELPKMNFTINRNTGCVISVTTGLDRESFPYTYLVQA 343

QY 324 PYLAHQWVNMWDVSVQLANHTVTKCKALGYSFAGTITD-----VSGINTQ 373
DB 344 ADLOG-EGISTATAV---ITVDTNDNPI---FNPITYKGVPENEANVTTLTKVD 396
QY 374 GTTEGDFWDLRLDOLGKPSSEIYECORP-----KPIILHSGELTIPHPQPI 423
DB 397 ADAPNTPAWEAVY-TILNDGGGOFVYTPVNNNDGLTKTAKGLDFEAKOOYLHVAVTN 455
QY 424 VDVOITVGSLSIAIPGELTITMSGRF-----REAIKKEFALYCKMDTVVJAGLSNY 478
DB 456 VPEEVSLETTSTATVY-DVLVDNEGPIFAPPEKREVSSEDFG--GGLTSTYAGPDPF 512
QY 479 THYITTEYEQOQRYAASTIYGPRT--LSATIQLEFRD-----LAKAIATDVAN 526
DB 513 MEOKITRYIRWTRNMLE--INPDGAISTRALREDREHVKNSYTAIITADNGSP 569
QY 527 MSSGPEPEFKMLIASLIPNIDRAPIGKHFQDVLQAPKREYGVGEVIEFGANPK-- 584
DB 570 VARG-----TGTLLILSDVNDNAPI-----PEPRT-----IFCERNPKFO 606
QY 585 ---NSAENQHOTELTVEKEDSVADMOIMYNDASMET 619
DB 607 VINHDADLPMTSPFTALHTGGRVPMNTIOYNDPTQES 645

RESULT 12
US-09-324-867-2
Sequence 2, Application US/09324867A
Patent No. 6251632
GENERAL INFORMATION:
APPLICANT: Lilliecap, David
APPLICANT: Cameron, Cherie
APPLICANT: No. 6251632ley, Colleen
APPLICANT: Horrocks, L. Suzanne Hoyle
APPLICANT: Hough, Christine
TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use
FILE REFERENCE: 1669.0010002/JAG/BJD
CURRENT APPLICATION NUMBER: US/09/324, 867A
CURRENT FILING DATE: 1999-06-03
EARLIER APPLICATION NUMBER: 09/035,141
EARLIER FILING DATE: 1998-03-059
EARLIER APPLICATION NUMBER: 60/039,953
EARLIER FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 63
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 2343
TYPE: PRT
ORGANISM: Canis familiaris
US-09-324-867-2

Query Match 2.9%; Score 104; DB 4; Length 2343;
Best Local Similarity 17.2%; Pred. No. 1.4;
Matches 144; Conservative 102; Mismatches 295; Indels 296; Gaps 37;
QY 87 RDNVILSAHTHSGRPGFYQTYLIASE---GFSNRTFYIYSGIMKSIDIANT---- 138
DB 1214 QENVALPQAHMTIGIKRNLK-NLFLSTKONVAGLEEQPYTILQDTRSLNDSPHSEGIH 1272
QY 139 -----NLKKGKIKINKGNVANOINSPSSYLLNPOSERARYSNNDKEVL--- 184
DB 1273 MANFSKIREANLE-----GLGNOTNOWERPST-----TRMSSNSQOHITOR 1317
QY 185 -----VLKLVLDNGEDGLLSWFAI 204
DB 1318 GKRLKQRLSGGEIKFERKVIANDSTQWKMNMNYLAGCTLQLEYNKEKRAITQSL 1377
QY 205 HGVSMNN-----SNHVSNDNMGYAA----- 225
DB 1378 SDCSMKNHVTIOMNSALPVAKESAPSVRHTDLTKIPSOH--NSSHLPASACNVTFRER 1435
QY 226 -----YLFEQEKNK-----GYLPGGPFVAGPASSNLGVSPN----- 258

DB 1436 TSGVQSGSHFLQEAQRNNLSLAEVLTIGTEGOKF-----SLGKSATNOPMYKLENT 1489
QY 259 -ILGPHCVMTGES-----CDNDKSTPCNGSPSCMASGPGODFESTHIIGRIYOK 309
DB 1490 VILQPOLSETSDKVELLSQVNHQDEDSF-----PRTSNDSGH-----LDLMGKTFLOK 1539
QY 310 AK-----ELYASAOEVTGPVLAHQVNMNTDVS-----OL 341
DB 1540 TCGPVAMKNTNSPGKPFPLKWTAESEKIPSKILGLAVMDNHNDQIPSEEWKOKKSQT 1599
QY 342 NATHYTKCKPMLGYSFAAGTIDVSGLNTITGTEGDFWDLRLQOLGKPSSEIYECOR 401
DB 1600 NTAFKRKDTILPLG---PCENDSTAINEGDKDPREAWM-----AKGEGRGLCS 1648
QY 402 KKPILL-HSGELTIPHPQPIVDVOITVGSLSIAIA---IPGELTMSGRFRFAIK 456
DB 1649 QNRPVSKNHQRETTV-TIQPEBDRKYEDDTSEIKKRDPIDYGYENQGLRSQOKTR 1707
QY 457 KEPAL-----YGMKMTVTI---AGLSNVYTHYITTEYEQOQRYEASTIYGPHTLS 506
DB 1708 HYFIAVERLWMDYGMKSRPHILRNRAOSGDVOQFKKVVQEFETDSF--TOPLYRGE-LN 1764
QY 507 AYIQLEFRDLAKAIATD---TVANNSSGPEPEFKMLIASLIPNIDRAPIDK----- 555
DB 1765 EHLGLQPYTRAEVEDNIVYTERKNOASRPY-SFYSSLISYDEDEGOGAPRRKFPVNPNET 1823
QY 556 --HFGDV--LOPAKPEY-----RVGEVVEYIFGANPKNSAENO--T 591
DB 1824 KIYFMKVOHNMAPTKEFDCKAMAYRSDVLDKDVHSGILGLTICRSTLTPAHRGROY 1883
QY 592 HOTELTVEKEDSVADMOIMY-----DASWETRFYWH-----KGI 627
DB 1884 VOEFALVFTIPETKSWYFTEMLERNCRAPCVOKEDPLKCNFRHAINGVKDTLPGL 1943
QY 628 LGLSNATYWHIPDT-AVPGYIRYRFGH--NRKOELLKPVVLAIFEGISSPFEV 680
DB 1944 VAAQDKVAMVYLLSMGSNENIHSHFSGHVFTYRKKEEYKAMVNYLIPGV---FETV 1997

RESULT 13
US-08-989-299-12
Sequence 12, Application US/08989299
Patent No. 6194556
GENERAL INFORMATION:
APPLICANT: Acton, Susan L.
APPLICANT: Robinson, Keith E.
TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/989,299
FILING DATE: 11-DEC-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold E., Beth
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: MIA-025.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 907 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-989-299-12

Query Match 2.8%; Score 100.5; DB 4; Length 907;
Best Local Similarity 18.7%; Pred. No. 0.58; Indels 219; Gaps 29;
Matches 116; Conservative 72; Mismatches 213;

134 DIAHT--NLKPKGI-----FINKGVANVOINRSPSYLLNPOSRRARYSN-- 178
107 DVAFSTAATPDELTEVIEIOLVDTFLNTGSIASNKTKGP--VFANPVAQALVNSSYWM 164
179 -----TKREMLVKLVLDNGE-----DLGLISWFAIHVVS-----MNN 211
165 KTDMLQAPGISIKDEKLRSLAGYEAEIKVLRVALSGWRYFNDASPSLKALDEAENV 224
212 SNHFNVDNMGAYALFEQE-----KNGYLPGGPFYAGFASMLGDSVSPNIIIGPH 263
225 LTFEVRSTSM--QAKQFDMASVTDEKYNRQLGVVSFEQ--MSALAPSHFADYSQ--AQA 277
264 CVNTGESQDNKSTCPNGSPCMASGPGQD--MESTHIIIRIITY-----QAKEL 313
278 ALNND-----SKDSTICDMDVPPCALQKIDMDSIFRNEKDSRLQHLWVSYTALAKSPS 334
314 Y-----ASASQEVTEGVLAHGVNMTDVSVOLNAT-----H 345
335 YNNITITISNEGAKINGFANGAMWRSASFDMSSKVKAEFDLNKOIKIYSTIOPFYQLH 394
346 TVKCKPAGLGSFAG--TIOGVSGLNITQGTGEDPMDLIDQLAKPSEIYECCQPK 404
395 AYMRQLAGTISNPGVSKDGPIDPAHLGSLDGD--WSAHYEQT--KPFEE--ESETPE 448
405 PILHSGELTIPHPQPDIVVOQIVTVGSLAIAIPGELTMSGRFREFA----- 454
449 AML-----SAFTQNTYTKKMFVTAY-----RYEFSAFPHLPKSYW 485
455 IKKEFALGKMDTVVLAG-----LSNVYTHYITTYEYEQAR 492
486 TTSIFARWYSKDMCHPAALDMRAPNDFRYKACAOJGEPDEQASHLIVQTYOY--- 541
493 YEASTLYGPTLSAYIQLPRDLAKAIAFDIVAN---MSSGPEPFEPKXNLIASIIPNIA 548
542 -----LTKDQSL-----LFEQASPVITDAIANAFALHSTNPHLYLSQKLVSEHLDIK 590
549 DRAPIGHFGVDLP-----AKPEYRGEVEVEIVFGANPKRISAEQTHQTFLEVEK 600
591 DSVIINKLKYKESLESFRLKPTTIAADNMYR-----ELFDGTVPKPK----- 631
601 YEDSVADWQIMYNDASWETR 620
632 -----LNDRWWEIR 640

RESULT 14
US-09-026-001A-10
Sequence 10, Application US/09026001A
Patent No. 6413760
GENERAL INFORMATION:
APPLICANT: Boonhoo, Amechand
APPLICANT: Seehra, Jasbir
APPLICANT: Shaw, Gray
APPLICANT: Sako, Dianne
TITLE OF INVENTION: HIGHLY PURIFIED MOCARRAGIN, A COBRA VENOM
TITLE OF INVENTION: THERAPEUTIC USES THEREOF
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge

STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/09/026, 001A
APPLICATION NUMBER: US/09/026, 001A
FILING DATE: 18-FEB-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15293B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-026-001A-10

Query Match 2.8%; Score 100; DB 4; Length 613;
Best Local Similarity 23.7%; Pred. No. 0.32;
Matches 54; Conservative 33; Mismatches 103; Indels 38; Gaps 10;

85 YRRDNVIL-----SAIHHSRPGAFQOTLYILIASSEGSNRTFOYISGIMKSIDI----- 135
221 YKRQKPVYKRRYEMINIMNMYNRLNHLALIGLEIWSNNEINVOVDVQATLDLGEW 280
136 AHTNLKPKGIKIPINKGVANVOINRSPSY-----LNPQSPRA--RYSSWTDEMLVLK 187
281 REKLLPKRKRDNOGLTIDFGKTPVGLAYIGSICNKSXSAVAVQDSST--SMAIT 338
188 LVLDNGEDLGLISWFAIH--PVSMMNSNHFVNSDNMGYALFEQ--EKNKGYLPQG 241
339 MAHENGHMG-----IHHDGPGCTCGSNKCMSTRTREPAYQFSSCSVRHQLLDR 392
242 PFVAGFASNSLGDVSPNIIIGPHCVNTGESQDNKSTCPNGSPCMAS 289
393 POCILNKLPLSTDIVSPICGNFVEYGECD-----C--GSPADCQA 433

RESULT 15
US-09-026-001A-18
Sequence 18, Application US/09026001A
Patent No. 6413760
GENERAL INFORMATION:
APPLICANT: Boonhoo, Amechand
APPLICANT: Seehra, Jasbir
APPLICANT: Shaw, Gray
APPLICANT: Sako, Dianne
TITLE OF INVENTION: HIGHLY PURIFIED MOCARRAGIN, A COBRA VENOM
TITLE OF INVENTION: THERAPEUTIC USES THEREOF
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/026,001A
 FILING DATE: 18-FEB-1998
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: BROWN, SCOTT A.
 REGISTRATION NUMBER: 32,724
 REFERENCE/DOCKET NUMBER: G152938
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 498-8224
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 621 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-026-001A-18

Query Match 2.8%; Score 100; DB 4; Length 621;
 Best Local Similarity 23.7%; Pred. No. 0.32;
 Matches 54; Conservative 33; Mismatches 103; Indels 38; Gaps 10;

OY	85 YRDNVIL-----SAITHSGPAGFFQYTLIILASEGFSNRTFOYIVSGIMKSIDT-----135
DB	229 YKDRKVIKRRVYEMINTMNMVYRLNFIHALIGLEIWSNRNEINVOQDVOATLDFGEW 288
OY	136 AHTNLRPGKIFINKGVANVQINRSPSY-----LLNPOSERA---RYSSNTDKEMLYLK 187
DB	289 REKKLLPKRRNDNAQLTGIDKGTIPVGLAYIGSICNPKSSVAVVODYSRT--SWVAIT 346
OY	188 LVDLNGEDGLISWFAIH---PYSMNSNHFNVSNDNMGYAAYLFEQ---EKNKGYLPGOG 241
DB	347 MAHEMGHNG-----IHHDGPSCCTCGSNKVCWSTRTERPAVOFFSSCSVREHQEYLLRDR 400
OY	242 PIVAGFASNLGDVSPNITLGPICVNTGESCNDNDKSTCPNGGPGSMCMAS 289
DB	401 POCILNKPPLSTDIVSPPIGNNFVEVGECD---C--GSPADCOSA 441

Search completed: July 3, 2003, 12:53:08
 Job time : 30 secs

5'143841
 4'997,760

GenCore version 5.1.6.
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 3, 2003, 12:39:23 ; Search time 73 Seconds

(without alignments)
1244.889 Million cell updates/sec

Title: US-09-937-521-14

Perfect score: 3582
Sequence: 1 FSGYIGVGRADCTGVSDI.....KPAVLAEGISPEFVVT 682

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1931.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3582	100.0	682	21	AAB33292	Mouse mature neutr
2	3582	100.0	756	21	AAB33291	Mouse neutral/alka
3	3312	92.5	745	22	AAB97029	Rat protein involy
4	3054	85.3	761	23	AAE13544	Human ceramidase p
5	1509.5	42.1	704	22	ABE58152	Drosophila melano
6	1090.5	30.4	646	21	AAV57307	P. aeruginosa matu
7	1090.5	30.4	670	21	AAV57314	P. aeruginosa cera
8	879	24.5	280	20	AA138464	Human secreted pro
9	167	4.7	35	21	AAB33290	Mouse neutral/alka
10	148	4.1	80	21	AA603568	Human secreted pro

11	146.5	4.1	1565	18	AAW18305	Photobadus lumin
12	131.5	3.7	1565	19	AAW56568	Toxin rcoB, encode
13	127	3.5	27	20	AAV38468	Human secreted pro
14	125	3.5	26	20	AAV38466	Human secreted pro
15	118.5	3.3	2013	23	ABBA4751	Listeria monocytog
16	116.5	3.3	766	22	AAU34979	Enterococcus faeca
17	116	3.2	1339	21	AAB42887	Human ORF2651
18	116	3.2	1355	22	AAW1859	Human polypeptide
19	113	3.2	1981	19	AAW42634	Protein sequence t
20	112	3.1	812	22	AAW40932	Human polypeptide
21	111.5	3.1	627	21	AAV95538	Bipolaris spicifer
22	111.5	3.1	627	21	AAV96762	Bipolaris spicifer
23	111.5	3.1	627	23	AAO18211	B spicifera phenol
24	111.5	3.1	952	22	AAW78715	Human protein sequ
25	109.5	3.1	582	20	AAV23920	Amino acid sequenc
26	109.5	3.1	699	20	AAV08471	F. balustinum CP70
27	109	3.0	629	22	AAU32412	Novel human secret
28	109	3.0	752	22	AAW39146	Human polypeptide
29	109	3.0	1978	20	AAW27230	Amino acid sequenc
30	107.5	3.0	937	22	AAW78714	Human protein sequ
31	106.5	3.0	2628	17	AAW6030	P. gingivalis haem
32	106.5	3.0	2628	19	AAW69488	Haemagglutinin pro
33	106	3.0	1021	7	AAW61316	Infectious durcal
34	104	2.9	878	15	AAW55060	Sequence of human
35	104	2.9	878	16	AAW85487	Human E-cadherin p
36	104	2.9	878	23	ABW81475	Human E-cadherin p
37	104	2.9	1352	22	ABG08357	Novel human diagno
38	104	2.9	2343	20	AAW80989	Canine factor VIII
39	104	2.9	2343	21	AAV57846	Canine Factor VIII
40	103.5	2.9	914	22	ABW66394	Drosophila melano
41	103.5	2.9	1026	22	ABW61775	H. pylori secreted
42	103	2.9	1413	18	AAW20725	H. pylori secreted
43	102.5	2.9	1026	21	AAW83025	Staufen protein of
44	101.5	2.8	865	22	AAW95798	Human protein sequ
45	101.5	2.8	865	22	AAW67610	Amino acid sequenc

ALIGNMENTS

RESULT 1	
AAB33292	standard; Protein: 682 AA.
ID	
XX	
AC	AAB33292;
DT	02-FEB-2001 (first entry)
XX	
XX	Mouse mature neutral/alkaline ceramidase, SEQ ID NO:14.
DE	
XX	
KW	Neutral/alkaline ceramidase; mouse; murine; recombinant production;
KW	cellular ceramide content control; antibody; membrane synthesis;
KW	lipid engineering; ceramide metabolism; drug development;
KW	mature ceramidase;
OS	Mus sp.
XX	
PN	W0200058448-A1.
XX	
PD	05-OCT-2000.
XX	
PF	24-MAR-2000; 2000WO-JP01802.
XX	
PR	26-MAR-1999; 99JP-0084743.
PA	(TAKI) TAKARA SHUZO CO LTD.
XX	
PI	Ito M;
XX	
DR	WPI: 2000-619079/59.
DR	N-PSDB: AA97634.
XX	
PT	Mammalian neutral/alkaline ceramidase applicable as lipid engineering

PT reagent for studying structure and functions of ceramide as well as
PT developing drugs for diseases associated with ceramide metabolism
XX
PS Claim 1: Page 62-65; 76pp; Japanese.

CC The invention relates to a mouse neutral/alkaline ceramidase and to
CC nucleic acids encoding it. The invention also relates to expression
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
CC ceramidase, the recombinant production of the ceramidase, an antibody
CC against the ceramidase, and a method for controlling the ceramide
CC content in cells and/or tissues by introducing the gene or its antisense
CC nucleic acid into the cells and/or tissues. The ceramidase can be used
CC as a lipid engineering reagent for studying structure and functions of
CC ceramide as well as developing drugs for diseases associated with
CC ceramide metabolism. The present sequence represents the mature mouse
CC neutral/alkaline ceramidase.

XX
XX
SQ Sequence 682 AA:

Query Match 100.0%; Score 3582; DB 21; Length 682;
Best Local Similarity 100.0%; Pred. No. 6.8e-313;
Matches 682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSGYITGVGRADCTGVSDINLMGYKNGONAGLLTRLSRAFIADPDGSRMAFVS 60
DB 1 FSGYITGVGRADCTGVSDINLMGYKNGONAGLLTRLSRAFIADPDGSRMAFVS 60

QY 61 ELCMISQRLRELVKRLKESKYGSLYRRDNYIIAITHSGPAGFPQYTYIIASEGFSNR 120
DB 61 ELCMISQRLRELVKRLKESKYGSLYRRDNYIIAITHSGPAGFPQYTYIIASEGFSNR 120

QY 121 TFQYIVSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSND 180
DB 121 TFQYIVSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSND 180

QY 121 TFQYIVSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSND 180
DB 121 TFQYIVSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSND 180

QY 181 KEMLVKLVLDNGEDGLISWFAIHVSNMNSHFNPSNDMGYAAVLFDEKKKGYLPQ 240
DB 181 KEMLVKLVLDNGEDGLISWFAIHVSNMNSHFNPSNDMGYAAVLFDEKKKGYLPQ 240

QY 241 GPFVAGFASNLGDVSPNLTGPHCVNTGSCDNKSTCPNGGSPKMGASGPGDMEESTH 300
DB 241 GPFVAGFASNLGDVSPNLTGPHCVNTGSCDNKSTCPNGGSPKMGASGPGDMEESTH 300

QY 301 IIGRIIYOKAKELIYASASQEVTPVLAHQWVNMVDVSVOLNATHVTCKPALGYSFAA 360
DB 301 IIGRIIYOKAKELIYASASQEVTPVLAHQWVNMVDVSVOLNATHVTCKPALGYSFAA 360

QY 361 GTIDVSGNLITOGTEGDPFMDTLRDOLGKPSSEIVECQKRPILLHSGELTTPHPQ 420
DB 361 GTIDVSGNLITOGTEGDPFMDTLRDOLGKPSSEIVECQKRPILLHSGELTTPHPQ 420

QY 421 PDIVDQIYTVGSLAIAIPGELTTMSGRFREAIKKEFALYKMDKTVIAGLSNVTYH 480
DB 421 PDIVDQIYTVGSLAIAIPGELTTMSGRFREAIKKEFALYKMDKTVIAGLSNVTYH 480

QY 481 YTTTVEYQAOQREASTIYGPHTLSAYIQLFRLDLAKAIATDVTANMSSGPEPPFKNLI 540
DB 481 YTTTVEYQAOQREASTIYGPHTLSAYIQLFRLDLAKAIATDVTANMSSGPEPPFKNLI 540

QY 541 ASLIPDIADRAPIGKHFQDVLOPAKPEYVGEVEVEIFGANKNSAENOTHTFTLVK 600
DB 541 ASLIPDIADRAPIGKHFQDVLOPAKPEYVGEVEVEIFGANKNSAENOTHTFTLVK 600

QY 601 YEDSVADQIMYNDAWETRFYWHKGIILSNATIYWHIPDTAIPGIYIRYEGHNRKOE 660
DB 601 YEDSVADQIMYNDAWETRFYWHKGIILSNATIYWHIPDTAIPGIYIRYEGHNRKOE 660

QY 661 LKPAVILAFEGISSPEFVVT 682
DB 661 LKPAVILAFEGISSPEFVVT 682

RESULT 2

AAB23291
ID. AAB23291 standard; Protein; 756 AA.

XX
AC AAB23291;

DT 02-FEB-2001 (first entry)

XX Mouse neutral/alkaline ceramidase, SEQ ID NO:13.

XX Neutral/alkaline ceramidase; mouse; murine; recombinant production;

KW cellular ceramide content control; antibody; membrane synthesis;

KW lipid engineering; ceramide metabolism; drug development.

OS Mus sp.

PN W0200058448-A1.

PD 05-OCT-2000.

PF 24-MAR-2000; 2000WC-JP01802.

PR 26-MAR-1999; 99JP-0084743.

PA (TAKI) TAKARA SHUZO CO LTD.

PI Ito M.

DR WPI: 2000-619079/59.

DR N-PSDB: AAA97633.

PT Mammalian neutral/alkaline ceramidase applicable as lipid engineering

PT reagent for studying structure and functions of ceramide as well as

PT developing drugs for diseases associated with ceramide metabolism

PS Example 4; Page 58-62; 76pp; Japanese.

CC The invention relates to a mouse neutral/alkaline ceramidase and to
CC nucleic acids encoding it. The invention also relates to expression
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
CC ceramidase, the recombinant production of the ceramidase, an antibody
CC against the ceramidase, and a method for controlling the ceramide
CC content in cells and/or tissues by introducing the gene or its antisense
CC nucleic acid into the cells and/or tissues. The ceramidase can be used
CC as a lipid engineering reagent for studying structure and functions of
CC ceramide as well as developing drugs for diseases associated with
CC ceramide metabolism. The present sequence represents the mouse
CC neutral/alkaline ceramidase.

XX
XX
SQ Sequence 756 AA:

Query Match 100.0%; Score 3582; DB 21; Length 756;
Best Local Similarity 100.0%; Pred. No. 8.1e-313;
Matches 682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSGYITGVGRADCTGVSDINLMGYKNGONAGLLTRLSRAFIADPDGSRMAFVS 60
DB 1 FSGYITGVGRADCTGVSDINLMGYKNGONAGLLTRLSRAFIADPDGSRMAFVS 60

QY 75 FSGYITGVGRADCTGVSDINLMGYKNGONAGLLTRLSRAFIADPDGSRMAFVS 134
DB 75 FSGYITGVGRADCTGVSDINLMGYKNGONAGLLTRLSRAFIADPDGSRMAFVS 134

QY 61 ELCMISQRLRELVKRLKESKYGSLYRRDNYIIAITHSGPAGFPQYTYIIASEGFSNR 120
DB 61 ELCMISQRLRELVKRLKESKYGSLYRRDNYIIAITHSGPAGFPQYTYIIASEGFSNR 120

QY 121 TFQYIVSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSND 180
DB 121 TFQYIVSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSND 180

QY 181 KEMLVKLVLDNGEDGLISWFAIHVSNMNSHFNPSNDMGYAAVLFDEKKKGYLPQ 240
DB 181 KEMLVKLVLDNGEDGLISWFAIHVSNMNSHFNPSNDMGYAAVLFDEKKKGYLPQ 240

QY 241 GPFVAGFASNLGDVSPNLTGPHCVNTGSCDNKSTCPNGGSPKMGASGPGDMEESTH 300
DB 241 GPFVAGFASNLGDVSPNLTGPHCVNTGSCDNKSTCPNGGSPKMGASGPGDMEESTH 300

QY 301 IIGRIYQAKELASASQEVYTGPIVLAHQVNMNTDVSVOLNATHVTCTCKPALGYSFAA 360
 DB 375 IIGRIYQAKELASASQEVYTGPIVLAHQVNMNTDVSVOLNATHVTCTCKPALGYSFAA 434
 QY 361 GTIDVSGSLNTTQGTBSPDPMFDTLRDQLCKPSEELIECOKPKPILHSGELTTPHPWQ 420
 DB 435 GTIDVSGSLNTTQGTBSPDPMFDTLRDQLCKPSEELIECOKPKPILHSGELTTPHPWQ 494
 QY 421 PDIVDQIVTVGSLAIAIPEGLTMSGRFRREAIKKEFALYGMKDMVTVAGLSNYTH 480
 DB 495 PDIVDQIVTVGSLAIAIPEGLTMSGRFRREAIKKEFALYGMKDMVTVAGLSNYTH 554
 QY 481 YITTEEYOQOAEYEAASYTGPHLSAVIQLEFRLAKAIAATDYANNSSGPEPPFKNLI 540
 DB 555 YITTEEYOQOAEYEAASYTGPHLSAVIQLEFRLAKAIAATDYANNSSGPEPPFKNLI 614
 QY 541 ASLIPNADRAPICKHFGDVLQPAKPEYRGEVEVEYFVGANPKNSAENQTHQFLVEK 600
 DB 615 ASLIPNADRAPICKHFGDVLQPAKPEYRGEVEVEYFVGANPKNSAENQTHQFLVEK 674
 QY 601 YEDSVADQIMYNDASMETREYMHKGLIGLSNATVHIIPDTAPGIRIRYFGHNKOE 660
 DB 675 YEDSVADQIMYNDASMETREYMHKGLIGLSNATVHIIPDTAPGIRIRYFGHNKOE 734
 QY 661 LKPAVILAFEGISSPEEVYTT 682
 DB 735 LKPAVILAFEGISSPEEVYTT 756
 RESULT 3
 AAB97029 standard; protein: 745 AA.
 XX ID AAB97029 standard; protein: 745 AA.
 AC AAB97029;
 XX 20-JUL-2001 (first entry)
 DE Rat protein involved in hydrolysis of the acid amide bond in ceramide.
 XX Rat; acid amide bond hydrolysis; ceramide; cytostatic; cancer.
 OS Rattus norvegicus.
 XX PNTJCP200105/890.FA
 XX 06-MAR-2001
 XX 23-AUG-1999 99JP-0235218.
 XX 23-AUG-1999 99JP-0235218.
 XX (MITU) MITSUBISHI CHEM CORP.
 XX DR MPI: 2001-304133/32.
 DE N-PSDB: AAF99977.
 PT Novel protein, used to treat abnormal cell growth e.g. in cancers, is
 capable of hydrolyzing the acid amide bond between sphingosine and the
 fatty acid in ceramide.
 PS Claim 1; Page 12-13; 14pp; Japanese.
 CC The present sequence is provided in a specification relating to a
 protein having a 761 residue amino acid sequence or its variant
 comprising a replacement, deletion, insertion, addition or reversion of
 at least one amino acid, but retaining the activity of hydrolyzing the
 acid amide bond between sphingosine and the fatty acid in a ceramide.
 CC The protein can be used to treat diseases caused by abnormal cell
 growth such as cancers.
 XX Sequence 745 AA;
 SQ

Query Match 92.5%; Score 3112; DB 22; Length 745;
 Best Local Similarity 91.8%; Pred. No. 1.6e-288;
 Matches 626; Conservative 25; Mismatches 15; Indels 16; Gaps 1;
 QY 1 FSGYIGYGRADCTGQVSDINLMGKNGQNAQGLTRLEFRATILADPDGSSNMAFVS 60
 DB 80 FSGYIGYGRADCTGQVSDINLMGKNGQNAQGLTRLEFRATILADPDGSSNMAFVS 139
 QY 61 ELCMISQRLREVLKRESKYGSILYRRPNVTLSTHHSFGAGFFQYTLTLASEGSNR 120
 DB 140 ELCMISQRLREVLKRESKYGSILYRRPNVTLSTHHSFGAGFFQYTLTLASEGSNR 199
 QY 121 TFQYIVSGIMKSIDIAHTNLKPKIFIKGNVAVQIRSSSYLNPQSERARVSSNTD 180-
 DB 200 TFQYIVSGIVASIDIAHTNLKPKAVLKNQVAVQIRSSSYLNPQSERARVSSNTD 259
 QY 181 KEMVLKLVLDNGEDGLISWFAIHPVSMNSNHFPVSDNNGVAAVLEQEKNGYLPQ 240
 DB 260 KEMVLKLVLDNGEDGLISWFAIHPVSMNSNHFPVSDNNGVAAVLEQEKNGYLPQ 319
 QY 241 GPVAGFASNLGQVSPNIIIGPHCVNCGESCDNDKSCPCNCGSPCMASGPGDMFESTH 300
 DB 320 GPVAGFASNLGQVSPNIIIGPHCVNCGESCDNDKSCPCNCGSPCMASGPGDMFESTH 379
 QY 301 IIGRIYQAKELASASQEVYTGPIVLAHQVNMNTDVSVOLNATHVTCTCKPALGYSFAA 360
 DB 380 IIGRIYQAKELASASQEVYTGPIVLAHQVNMNTDVSVOLNATHVTCTCKPALGYSFAA 439
 QY 361 GTIDVSGSLNTTQGTBSPDPMFDTLRDQLCKPSEELIECOKPKPILHSGELTTPHPWQ 420
 DB 440 GTIDVSGSLNTTQGTBSPDPMFDTLRDQLCKPSEELIECOKPKPILHSGELTTPHPWQ 499
 QY 421 PDIVDQIVTVGSLAIAIPEGLTMSGRFRREAIKKEFALYGMKDMVTVAGLSNYTH 480
 DB 500 PDIVDQIVTVGSLAIAIPEGLTMSGRFRREAIKKEFALYGMKDMVTVAGLSNYTH 559
 QY 481 YITTEEYOQOAEYEAASYTGPHLSAVIQLEFRLAKAIAATDYANNSSGPEPPFKNLI 540
 DB 560 YITTEEYOQOAEYEAASYTGPHLSAVIQLEFRLAKAIAATDYANNSSGPEPPFKNLI 619
 QY 541 ASLIPNADRAPICKHFGDVLQPAKPEYRGEVEVEYFVGANPKNSAENQTHQFLVEK 600
 DB 620 ASLIPNADRAPICKHFGDVLQPAKPEYRGEVEVEYFVGANPKNSAENQTHQFLVEK 663
 QY 601 YEDSVADQIMYNDASMETREYMHKGLIGLSNATVHIIPDTAPGIRIRYFGHNKOE 660
 DB 664 YEDSVADQIMYNDASMETREYMHKGLIGLSNATVHIIPDTAPGIRIRYFGHNKOE 723
 QY 661 LKPAVILAFEGISSPEEVYTT 682
 DB 724 LKPAVILAFEGISSPEEVYTT 745
 RESULT 4
 AAE13544 standard; protein: 761 AA.
 XX ID AAE13544 standard; protein: 761 AA.
 AC AAE13544;
 XX 26-FEB-2002 (first entry)
 DE Human ceramidase protein.
 XX Human; ceramidase; nephrotropic; antipsoriatic; antitense therapy;
 KW gene therapy; proliferative disorder; cancer; cardiovascular disease;
 KW inflammation; neurodegenerative disorder; cytostatic; immunosuppressive;
 KW nootropic; signal transduction; breast cancer; autoimmune disorder;
 KW Alzheimer's disease; growth deficiency; lesion; lupus nephritis;
 KW glomerular disease.
 XX Homo sapiens.
 OS Homo sapiens.
 XX Location/Qualifiers
 FH Key

[illegible]

Dd	80	FSGCHIGVAGADCTGVVAOINLMGKVGSGQANQGLITFLYSRAPIIMAPDGSNNTVYVSI	139
OY	61	ELCMISORLELVLEAKLESKYGSILYRDNVYILSAITHSHSGPAGFOYTLTYLASEGFSNR	120
Dd	140	DIGWVSORLELVLEANLRQSKYSLYRDNVILSGTHHTSHSGAGFYQVTVFEVASEGFSNQ	199
OY	121	TFQYIYSGIMKSLDIAHTNLKPGKIFINKGVANVQIRSRSSVLLMPQSRARYSSNTD	180
Dd	200	TFQHMWTGILKSLDIAHTNMKPGKIFINKGVNDVQIRSPYSTLQNDQSRARYSSNTD	259
OY	181	KEMVLVLYLVLDNEEDGLISMEFAIPVSNMNSNHFPVNSDNGVYAAAYLFEQEKNGYLPQ	240
Dd	260	KEMVLVLYLVLDNEEDGLISMEFAIPVSNMNSNHLYNSDNGVASYLLEQEKNGYLPQ	319
OY	241	GPEVAGFASSNLGDVSPNIIIGPHCVNTEGSCDNDKSTCPNGPSCMAASGPGDMEFSTH	300
Dd	320	GPEVAAFASSNLGDVSPNIIIGPHCVNTEGSCDNDKSTCPIGPSCMAKAGPGDMEFSTQ	379
OY	301	IIGRIIYQKKELIYASQSEVTGPVLAHQVYNNMTDVSVOIUNATHYVTKTCKPALGYSFAA	360
Dd	380	IIGRAMVQRAKELIYASQSEVTGPVLAHQVMDTDVYVNNSTHASKTCIPALGYSFAA	439
OY	361	GTTIDGYSGLNITPGTGGEPFMDTLRDOLLKPSSEIYECOKPKPILHSGELTIPHPWQ	420
Dd	440	GTTIDGYSGLNITPGTGGEPFMDTLRDOLLKPSSEIYECOKPKPILHSGELSKRPWH	499
OY	421	PDIVDVOIYVWGSIAIAIPGELITTMSGRRFREALIKKEFALYGMKDMTVYIAGLSNYTH	480
Dd	500	PDIVDVOIITLGLSLATTAIPGELITTMSGRRLEAVQAFAHGMQNMTVYISGLCNYTH	559
OY	481	YITTYEYQAKRYEASTIYGPHITLSAYIQLEFRLAKAIATDYVYANSSGPEPPEFNLI	540
Dd	560	YITTYEYQAKRYEASTIYGPHITLSAYIQLEFRLAKAIADTYVYANSSGPEPPEFKOLI	619
OY	541	ASLIPNADAPRIKGKHEGVDLQPAKPEYRVGVEVYEVIVGANGPKNSAENQTHOTFLVEX	600
Dd	620	VPLIPSTVDAPRGKTEGVDLQPAKPEYRVGVEVYEVIVGANGPKNSVQNOTHOTFLVEX	679
OY	601	YEDSVADMQIYNDASMETRFYWNHGIIGLSNATYVMIIPDTAARGIYRIYREGHNKQD	660
Dd	680	YEATSTSMQIYNDASMETRFYWNHGIIGLSNATYVMIIPDTAARGIYRIYREGHNKQD	739
OY	661	LKPAVILAEFGISSPEEVYT 681	
Dd	740	ILKPAVILSEFGTSPAFEVYT 760	
RESULT 5			
ABBS5152	ID	ABBS5152 standard; Protein; 704 AA.	
XX	ABBS5152:		
AC	XX	26-MAR-2002 (first entry)	
DT	XX	Drosophila melanogaster polyptide SEQ ID NO 1248.	
DE	XX	Drosophila: developmental biology; cell signalling; insecticide;	
KW	XX	pharmaceutical.	
XX	XX	Drosophila melanogaster.	
OS	XX	MO2001.71042-A2.	
PN	XX	27-SEP-2001.	
PD	XX	23-MAR-2001; 2001MO-US09231.	
XX	XX	23-MAR-2000; 2000US-191637P.	
XX	XX	23-MAR-2000; 2000US-0614150.	
XX	XX	(PEKE) PE CORP NY.	

PI	Venter JC, Adams M, Li PMD, Myers EM,
XX	
DR	WPI: 2001-656860/75.
DR	N-PSDB: ABL02255.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
PS	Disclosure: SEQ ID NO 1248; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL01840-ABL03511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pt_sequences .
XX	
50	Sequence 704 AA:

Query Match	42.1%;	Score 1509.5;	DB 22;	Length 704;
Best Local Similarity	44.6%;	Pred. No. 2.2e-126;		
Matches 311;	Conservative 114;	Mismatches 230;	Indels 43;	Gaps 14

```

QY      4 YTTGVRACCTGQVSDINIMGYKGNQGNARGLTFTLFERATILADPDSJNMAFVSJELC 63
Db      25 YKVGVRADITGPPEALINPMGYANIKQYGRGITHRTVFARATVDEDSG--NRVAFVSADAG 83

QY      64 MISQRLLEVLKRLSEKYGSLYRDVNLTLSAITHHSGPAGFYQYTYLLASEGFSNFTQ 123
Db      124 YIYSGIMKSIDIAHNLKPGKFEIFINKGVANVOJNRSPSYLLNPOJEBARAYSNTDKEM 183
QY      144 YMAQGLYLCIKRRYDLVDGRITLLSKTLYLVANNINRSPSYLLNPAERARQYEHDDIKTL 203
Db      184 LVLTLDNGEDIGLISWFAIPHPVSMNNSNHFPVSDNNGYAAYFEDEKKNKYLPQGGP 243
QY      204 TQLRPYDLLENLLGAFNMYVAHVAHSMNNTNRLVYSDNAGYALLLEKEYNPNKKMPGKGK 263
Db      244 VAGFASSNIGDYSPNLTGPHCYNIGESCNDNKSCTPCNGPSCMAASGPGQDMPESTHIIIG 303
QY      264 VQAFGCSNIGDVSYPNIMGKCSISQNECDLTSRCPT--GEEDCFCAASGPGKDMESTQILG 322
Db      304 RIYYKRA-----KELYASASQEVTPVPLAHOVMNTDVSQVLNAYHT-----VKTC 350
QY      323 QRLADALGLLNQOSESTARREYTDVAFHQPDMRYN---GSYTNPLSRKYDKIKGC 379
Db      351 KPALGYSFAAGTIDGVSGNLTQGTTEGDPFMDTLRDQLLKGPSBEIIVECOKPAILHS 410
QY      380 QPAMGYSFAAGTTDEGAFSFEQGTDTNPMNMFVRD--FIAPQEDIKCHEPILLAT 438
Db      411 GELTIHPMOPDIUVQIYTVGSLSAIAIPBELTTHMSRRRRLAKKEPALYCGKDMITY 470
QY      439 GRATPEYEWOPKIVSDQLKIGDVIIAVPEFTTMAGRRLRNOJRAAASVGVGDIETVI 498
Db      471 IAGLSVNYHYITTYEEDYQQRYEAASTYTPHTLSAIOLEFROIARAKAITDVANNSSG 530
QY      499 IAGLTINISYTYTPEBYOQRYEAASTIIPGHHHSIMDYFERJLTAMMRN--EIVYDAG 556
Db      531 PEPPFKNLIASLIPNIA--DRAPIGKHFGDVLQAPKPEYRVEGVEVYIYVANGKNSAEN 569
QY      557 PSEPPYNDMLSLNTGVLEFDGHPINTDGYKYSQPNKEYGJINELVAKTYISGNRNRL-- 614
Db      590 QTFQTFLEVEK--YEDSVADMOIMNDSWENRFPWHR--GILLSNATYIWHIPDFAVP 645
QY      615 FTKTYITTEKRIKNEDR---KVAAYTDSWETKAMVHAKTNLTILFSEMDIYWDISPTQLP 671

```

```

QY      646 GIYRIYFGINRKQELLKPAVIA---FEGISSPEV 679
      | | | | : | : | | |
Db      672 GEYRIHSGEYK-----YLGGKYPEGLTHSEFV 701

```

RESULT 6
AAV57307
ID AAV57307 standard; Protein; 646 AA
vv

AC AAY57307;
VY

DT 05-JUN-2000 (first entry)
 VY

P. aeruginosa mature ceramidase sequence

KW Ceramidase; ceramide; sphingosine; fatty acid; atopic dermatitis
KW lipid engineering; antidermatitis.

Pseudomonas aeruginosa

PN EP980912-A1
XY

PD 23-FEB-2000
YY

PF	20-AUG-1999;	99EP-0116154
YY		

PR 20-AUG-1998; 98JP-0234769
XX

PA (TAKI) TAKARA SHUZO CO LTD
XX

PI Okino N, Ito M;
XX

DR WP1; 2000-258590/23
DR N-PSDB: AA790566

XX
PT New Pseudomonas

XX diagnosis, treatment

PS CLAIM 1; Page 10-18; 32pp; English
XX

CC This represents a mature alkaline ceramidase from *Pseudomonas aeruginosa*.
CC Ceramidase can be produced by standard recombinant methodology.
CC Ceramidase converts ceramide to sphingosine and a fatty acid, causes, or
CC exacerbates, atopic dermatitis. The ceramidase gene, polypeptide, derived
CC oligonucleotides and ceramidase-specific antibodies are used to prevent
CC and/or treat atopic dermatitis. The gene and the polypeptide can be used
CC as reagents for lipid engineering and for analyzing the structure and
CC activity of ceramidases.

SQ Sequence 646 AA;

Query Match	30.48;	Score 1090.5;	DB 21;	Length 646
Best Local Similarity	35.98;	Pred. No. 9.8e-89;		

[illegible]

00: 64 MICROBI ENV KPI ECEVCGI VDDOWITI CATUHUCCDACEEAVETI VTI ACECECNBTEO 1221

```

65 MTEOAVHFKVIAPIKAKYDGVYDENNTYMAATHTHSCBGFESHYAMYNI.SVI.GEOEKTEN 124

```

124 YIVSGIMKSTDIATHTNPKDKTEFNKCNVANYOINBSPSSYIJNPOSEBARYSSNTDKEM 183

Db 125 ATVDGIVRSTERAQARIOPGRI.EYGSGET.RNASRNRSL.SHI.KNP--DIAGYEDGIDPOM 182

184 T.VI.KI.YDI.NGEDT.GITSWFATHPV.SMNSNH.FVNSDNMGY.AYT.FEOFKNKGYL.PGOGPF 243

Db 183 SVLSFVDANGELAGAISWFPVHSTSMTNANHLISPDNKGYASYHWEHDVSR-----KSGF 237

244 VAGFASSNLGDVSPNII LGPHCVNTGESCDNDKSTCPNGGPPSMCMASGPGODMEESTHIIG 303

[illegible]

Db 238 VAAFAQTNMGNSLNL-----NLKPGSGPPDNEFDNTRIG 273
QY 304 RIITYOKAKELIYASASQEVGTGTVLAHQWVMTDVSQVLAAT--HTVKTCKPALGYSFAAG 361
Db 274 LRQFAKAYETIAGQAQOEVEVLGELDSRFRFVDFTLPIRPEPTDGPQLCTAAIGTSLAAG 333
QY 362 -TIDVSGLNTIGCTEGDEPFMTLDQLLGKPSSEIVEQCKPKILLHSGELTIHPMQ 420
Db 334 STEDGCGPLGLEGG---NNPFSLALGGLTGVPQELVOCQAERTILLADGNKK-PYPMT 389
QY 421 PDIVDQIYVYVSGSLAIAIPGELTMSGRFRFAIKKEFALYGMKMTVYIAGLSNYTH 480
Db 390 PTVLPLOMERIGLDELGAPEFTVAGVIRRAVQAASAAIGIRH--VYFNQYANAYAS 447
QY 481 YITTYEYQAYRVEASTIYGPHTLSAYIQLFRDLAKA-----IATDVA-----N 526
Db 448 YVTTREYVAAQEEGSGTLGPTQAAVQOOLFVDMVALRERLPVETSALAPLSCQMN 507
QY 527 MSSGPPPPFKNLIALIPNIADRAPIGKHFGDVLQPAKPEYRVGEVVEYIEFGANPKNS 586
Db 508 FQTVG-----VADDPYIGKSGFDVLOQPRESYRIGDKVTVAFTGHPKN- 551
QY 587 AENQTHOTFLTEVEKY-EDSVADMQIMVNDASMETREFYMHKIGLSNATYWHIIPTAAP 645
Db 552 -DLRTEKTFLEVYVNIKDKGQTPVYATDNDMDTQYRWERVIGISAKATISMSIPGTER 610
QY 646 GIYRIRYFGHNRKQELLKPAVILAFEGISSPEFVVT 682
Db 611 GHYIRHYGNKAKNFWTK---ISEIGSGTSPFEVLGT 644

RESULT 7

AA57314
ID AA57314 standard; Protein; 670 AA.

XX AA57314:

DT 05-JUN-2000 (first entry)

XX P. aeruginosa ceramidase sequence.

XX Ceramidase; ceramide; sphingosine; fatty acid; atopic dermatitis;

KW lipid engineering; anti-dermatitis.

XX Pseudomonas aeruginosa.

XX BP980912-A1.

XX 23-FEB-2000.

XX 20-AUG-1999; 99EP-0116154.

XX 20-AUG-1998; 98JP-0234769.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Okino N, Ito M;

XX MPI: 2000-258590/23.

XX N-PSDB; AA290573.

XX New Pseudomonas aeruginosa alkaline ceramidase gene, useful for

XX diagnosis, treatment and prevention of atopic dermatitis -

XX Example 1; Page 24-27; 32pp; English.

XX This represents a ceramidase from Pseudomonas aeruginosa strain AN-17.

XX Ceramidase can be produced by standard recombinant methodology.
XX Ceramidase converts ceramide to sphingosine and a fatty acid, causes, or
XX exacerbates, atopic dermatitis. The ceramidase gene, polypeptide, derived
XX oligonucleotides and ceramidase-specific antibodies are used to prevent
XX and/or treat atopic dermatitis. The gene and the polypeptide can be used
XX as reagents for lipid engineering and for analyzing the structure and

CC activity of ceramidases.

XX Sequence 670 AA;

QY Query Match 30.4%; Score 1090.5; DB 21; Length 670;
Best Local Similarity 35.9%; Pred. No. 1e-88;
Matches 250; Conservative 118; Mismatches 254; Indels 75; Gaps 14;

QY 4 YIIGVRADCTGVSDINIMGYKNGCOMARGLTLRLFSRAFTLADPGSNMAFVSELC 63
Db 29 YRGLGKADITGAAEAGVMGYSLSBQKTGAGIHMRQMARAFVIEEASGRLLVYVMTDGL 88
QY 64 MISQRLRLVLALESKYSLYRDVYLLSAITHHSGPAGFPQYTYLILASBFSNRTQ 123
Db 89 MTFQAVHLKYLAKKAPVYVDENVYMLAATHTHSGPGGFSHYAYVNYLSVLGFEKTEPN 148
QY 124 YIYSGIMKSIDIHNTLTKPKIRFINKGNVANOINSPPSYLNPQSERARYSSNDKEM 183
Db 149 AIVDGIYRSTRERQARLQPRRLFYGGSELRLNBSRNSLSHLKNP--DIAGYEDGIDPQA 206
QY 184 LVKLVDLNGEDLGLISWFAIHPVSMNNSNHPVNSDNMGYAAVLFQEKNGYLPQOGAP 243
Db 207 SVLSFVDANGELAGAISWFEVHSTSMTNANHLISPDKGYASVHWEHVDYSR-----KSGP 261
QY 244 VAGFASSNLGDVSPNLTGPHCVNTGSCDNDKTCNGGSPKMGMAAGPQDMFEETHIG 303
Db 262 VAAFAQTNMGNSLNL-----NLKPGSGPPDNEFDNTRIG 297
QY 304 RIITYOKAKELIYASASQEVGTGTVLAHQWVMTDVSQVLAAT--HTVKTCKPALGYSFAAG 361
Db 298 LRQFAKAYETIAGQAQOEVEVLGELDSRFRFVDFTLPIRPEPTDGPQLCTAAIGTSLAAG 357
QY 362 -TIDVSGLNTIGCTEGDEPFMTLDQLLGKPSSEIVEQCKPKILLHSGELTIHPMQ 420
Db 358 STEDGCGPLGLEGG---NNPFSLALGGLTGVPQELVOCQAERTILLADGNKK-PYPMT 413
QY 421 PDIVDQIYVYVSGSLAIAIPGELTMSGRFRFAIKKEFALYGMKMTVYIAGLSNYTH 480
Db 414 PTVLPLOMERIGLDELGAPEFTVAGVIRRAVQAASAAIGIRH--VYFNQYANAYAS 471
QY 481 YITTYEYQAYRVEASTIYGPHTLSAYIQLFRDLAKA-----IATDVA-----N 526
Db 472 YVTTREYVAAQEEGSGTLGPTQAAVQOOLFVDMVALRERLPVETSALAPLSCQMN 531
QY 527 MSSGPPPPFKNLIALIPNIADRAPIGKHFGDVLQPAKPEYRVGEVVEYIEFGANPKNS 586
Db 532 FQTVG-----VADDPYIGKSGFDVLOQPRESYRIGDKVTVAFTGHPKN- 575
QY 587 AENQTHOTFLTEVEKY-EDSVADMQIMVNDASMETREFYMHKIGLSNATYWHIIPTAAP 645
Db 576 -DLRTEKTFLEVYVNIKDKGQTPVYATDNDMDTQYRWERVIGISAKATISMSIPGTER 634
QY 646 GIYRIRYFGHNRKQELLKPAVILAFEGISSPEFVVT 682
Db 635 GHYIRHYGNKAKNFWTK---ISEIGSGTSPFEVLGT 668

RESULT 8

AA38464
ID AA38464 standard; Protein; 280 AA.

XX AA38464:

XX 18-OCT-1999 (first entry)

XX Human secreted protein encoded by gene No. 21.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
XX diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;
XX developmental abnormality; foetal deficiency; blood; allergy; renal;
XX immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;
XX inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
XX cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;

XX	26-FEB-1999;	99US-0122487.
PR	(GEST) GENSET.	
XX	Dumas Milne Edwards J, Duclert A, Giordano J;	
XX	WP1; 2000-500381/45.	
DR	N-PSDB; AAC03574.	
XX	New nucleic acid that is a 5' expressed sequence tag (5' EST) for	
PT	obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for	
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -	
PS	Claim 13; SEQ ID 7649; 71pp + CD-ROM; English.	
XX		
CC	The present sequence is a polypeptide encoded by one of a large number	
CC	of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs	
CC	were prepared from total human RNAs or polyA+ RNAs derived from 30	
CC	different tissues. EST sequences usually correspond mainly to the 3'	
CC	untranslated region (UTR) of the mRNA because they are often obtained	
CC	from oligo-dT primed cDNA libraries. Such ESTs are not well suited for	
CC	isolating cDNA sequences derived from the 5' ends of mRNAs and even in	
CC	those cases where longer cDNA sequences have been obtained, the full 5'	
CC	UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'	
CC	ends and can therefore be used to obtain full length cDNAs and genomic	
CC	DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and	
CC	chromosome mapping procedures. They are used to obtain upstream	
CC	regulatory sequences and to design expression and secretion vectors.	
SO	Sequence 80 AA:	
Query Match	4.1%; Score 148; DB 21; Length 80;	
Best Local Similarity	35.7%; Pred. No. 1.9e-05;	
Matches 35; Conservative	9; Mismatches 22; Indels 32; Gaps 2	
OY	479 THYTTYEYQAGREASTIYGPTLSAYIDLFRLAKAINTDVIYANNSGPE----PP 534	
DB	11 THYLTFSS-----SEXLAKGYCYGHGSOPEQRSRYS 42	
OY	535 FFKNLIASLIPNIDRAPIGKHFQDVLPAKPEYRNGE 572	
DB	43 LFGTIIVLPISYVDRAPKGRFGDVLPAKPEYRNGE 80	
RESULT 11		
AAWI8305		
AC	AAWI8305 standard; Protein: 1565 AA.	
XX		
XX	AAWI8305;	
DT		
DT	30-JAN-1998 (first entry)	
DE		
XX	Photorhabdus luminescens Insect toxin TcCB.	
XX		
KW	Insecticide; Insect; pest control; biological control;	
KW	Photorhabdus luminescens; TcCB; Southern corn rootworm;	
KW	Colorado potato beetle; Western corn rootworm; meal worm;	
KW	boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;	
KW	cabbage looper; codling moth; corn earworm; European corn borer;	
KW	tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;	
KW	Diptera; Dictyoptera; Acarina; Homoptera.	
XX		
OS	Photorhabdus luminescens strain W-14 (ATCC 55397).	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..11
FT	/note= "N-terminal peptide (Claim 30)"	
PN	W09717432-A1.	
XX		
XX	15-MAY-1997.	
XX		

XX	06-NOV-1996:	96WO-USI8003.
XX	28-AUG-1996:	96US-0705484.
PR	06-NOV-1995:	95US-0007255.
PR	28-FEB-1996:	96US-0608423.
XX	(MISC) WISCONSIN ALUMNI RES FOUND.	
PA	Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;	
PI	Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GU,	
PI	Petrell J, Roberts JL, Rochelleau TA, Schoonover S;	
PI	Strickland JA;	
DR	WPI; 1997-281022/25.	
DR	N-PSSB; AAT68849.	
XX		
PT	Photornhabdus sp. insecticidal protein toxins and DNA encoding them -	
PT	can be genetically engineered into insect larvae food and plants for	
PT	insect control	
XX		
PS	Claim 34: Page 239-245; 276pp; English..	
CC	This polypeptide comprises the 165 kDa TcCB insecticidal toxin	
CC	protein of Photornhabdus luminescens W-14. Its sequence was deduced	
CC	from a genomic DNA clone (AA168849) and includes an isolated	
CC	N-terminal peptide (see AA178677). Claimed toxins of P. luminescens	
CC	(see AA178711, AA17884-89, AA17899-900, AA18301-06) can be produced	
CC	by recombinant DNA methods and applied to, or genetically engineered	
CC	into, insect larvae food and plants for insect control. The toxins	
CC	are particularly effective against Southern corn rootworm, Colorado	
CC	potato beetle, Western corn rootworm, meal worm, boll weevil and	
CC	turf grub (Coleoptera), beet armyworm, black cutworm, cabbage looper,	
CC	coddling moth, corn earworm, European corn borer, tobacco hornworm	
CC	and tobacco budworm (Lepidoptera), and are also active against	
CC	insects of the orders Hymenoptera, Diptera, Dictyoptera, Acarina	
CC	and Homoptera. (All claimed).	
XX		
SQ	Sequence 1565 AA:	
	Query Match 4.1%; Score 146.5; DB 18; Length 1565;	
	Best Local Similarity 20.9%; Pred. No. 0.0039;	
	Matches 144; Conservative 78; Mismatches 253; Indels 215; Gaps 33	
OY	37 TRLFSAFLAPDDSNKMAFVSVELCMISGRALBEVLKLSEKSGLYRNDVILSAH 96	
DB	656 YTTLSAEADSTDPDKN-----YLQVCL-----NWWDHYD 685	
OY	97 THSGPAGFEQ-----YTLYLASEGFSNRTOYIVS-----GIKSIDIAHTNLKPRI 145	
DB	686 RSQGKGKAASWSKKMFNYVALQDSKADAIPLRVSVYDSKRGLGYQLDPMWSSL--PAT 744	
OY	146 FINKGNV-----ANVOINRSBSSTLL--NPQSERARYSSNTDEKMLVKLVLDNGEDIG 197	
DB	745 RLITTEFVRRLIEKAMLGID-SILLDTYTLOADESLADLVITGKSBPV-----DFNGSN-G 796	
OY	198 LISMFALHPVSMNNNSNHVNSDNMGYA-----YLEEEO-FNKGLGQGFVFAGFASSML 252	
DB	797 LTFWELEFHLPLPVATRFRANEQSPSPAKSLIHITYFDPAKNK---PHNA-----AY 845	
OY	253 GDVSPNILEPHCVNTGESCDNDKSTCPNGSGPSCMASPGQDMFEETHIIIGRIIYOKA-- 310	
DB	846 MWVRPLVEG---NSDLSRHLDDSIDP-----DPOAVAH---PVYQKAVE 884	
OY	311 -----KELYASAQSEVNGP-----VLAHH 329	
DB	885 IAVVSNLIAQGDMMYROLTRDGLTQARAYTYULAEELLGPDPDVSLSSIWTPTQTLDTLAG 944	
OY	330 QMWNMVDVSVOLNATHVTCTCKPALGYSFAGATIDGV--SGLNITGGTEGDFEMDTLRD 387	
DB	945 OKRAVLBDPEHQIANDTDALRALPGNNVSYLKLANQGYFNELPLAVNLMSS-----HMOTJDA 999	
OY	388 QLL-----GRSEETVEQCKRPILL-----HSGETLTIRHPQOPRIDVQIVTVGS 433	

Db	1000	RLYMLRHNLVTGDKRPLSLPTLYAAVPDVALLAQAOSGTLT-----N;YSAGMLTFTGT	1052
Qy	434	L-AIAPAIP-----GELTIMG-----RRREAIKKEFALYCKMDT-----VYA	472
Db	1053	LTSFAMLPRAYSAVGTLTSPQONLLSLERSEERACOEELAAQOOLLDMSSVAITLQOQALD	1112
Qy	473	GLSNVYHYITVEEYQAOREAASTYIGPPTLSAVYIOLPFDLAKA-----IATDTVANMS	528
Db	1113	GLADRLALLASQATAQ-QRHDHYTLY-QNNISSAEQLVMDTQTSQSLISSSTGYOTA	1170
Qy	529	SGPEPPEFKMLIASLIPNADRAPIGKHFQDVLQPARPEYRVGEVEVIEFGANPKNSAE	588
Db	1171	SG-----ALKVIFNIFGLADGS-----RYEVEFELALGMAAQAT	1208
Qy	589	NOTHOTFLTYEKEYEDSVADQIMYNDSWE	618
Db	1209	SYVAERLATTEHYRRRREEMOIYOQAQSE	1238
RESULT 12			
AAW56568			
ID	AAW56568	standard; Protein; 1565 AA.	
XX	AAW56568;		
XX			
DT	07-AUG-1998	(first entry)	
XX			
DE	Toxin TccB, encoded by the tccb gene from genomic region tcc.		
XX			
KW	Photorhabdus luminescens W-14; nematode; symbiotic;		
KM	Heterorhabdilis; tcc; tcb; tcc; tcd; insecticidal activity; toxin;		
KW	Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;		
KM	Homoptera; Soltwey; Western corn rootworm; Colorado potato beetle;		
KW	mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;		
KM	cabbage looper; codling moth; corn earworm; European corn borer;		
KW	Tobacco hornworm; budworm.		
XX			
OS	Photorhabdus luminescens.		
XX			
XX	W09808932-A1.		
XX			
PD	05-MAR-1998.		
XX			
PF	05-MAY-1997; 97MO-US07657.		
XX			
PR	06-NOV-1996; 96MO-US18003.		
PR	28-AUG-1996; 96US-0705484.		
PR	06-NOV-1996; 96US-0743699.		
XX			
PA	(DOMC) DOWELANCO.		
PA	(WISC) WISCONSIN ALUMNI RES FOUND.		
XX			
PI	Blackburn MB, Bowen DJ, Cliche JA, Ensign JC, Fatig H;		
PI	Pfrench-Constant RH, Guo L, Hey TP, Merlo DJ, Orr GJ;		
PI	Petrell J, Roberts JL, Rocheleau TA, Schoonover S;		
PI	Strickland JA, Sukhapinda K;		
XX			
DR	WPI; 1998-179427/16.		
DR	N-PSDB; AAV29926.		
XX			
PT	Isolated toxins from Photorhabdus luminescens strains - useful for		
PT	control of insect pests		
XX			
XX	Claim 34; Pages 276-281; 321pp; English.		
CC	The present sequence represents a protein named TccB of the bacterium		
CC	Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the		
CC	nematodes of the Heterorhabdilis genus. The bacterium has at least 4		
CC	distinct genomic regions, tccA, tcb, tcc, and tcd. Peptide products are		
CC	produced from these regions that are associated with insecticidal		
CC	activity. The native toxins are secreted proteins. The proteins are		
CC	toxic to insects upon exposure and especially when ingested. The		
CC	nucleic acid sequence can be used to produce transgenic plants.		

KM Human; secreted protein; fusion protein; gene therapy; protein therapy;
KM diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KM developmental abnormality; foetal deficiency; blood; allergy; renal;
KM immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KM inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid digestion;
KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO935158-A1.
XX
PD 15-JUL-1999.
XX
PF 06-JAN-1999; 99WO-US00108.
XX
PR 07-JAN-1998; 98US-0070704.
PR 07-JAN-1998; 98US-0070657.
PR 07-JAN-1998; 98US-0070658.
PR 07-JAN-1998; 98US-0070692.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
XX
DR WPI: 1999-444190/37.
DR N-PSDB: AA206239.
XX
PT New isolated human genes and the secreted polypeptides they encode
XX
PS Disclosure; Page 215; 227pp; English.
XX
SS This sequence represents a secreted human protein encoded by the gene
CC clone detailed in the descriptor line. The gene can be used to generate
CC fusion proteins by linking to the gene to a human immunoglobulin Fc
CC portion (e.g. AA206210) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 36 novel genes and their fragments (nucleic
CC acid sequences: AA206219-206263; amino acid sequences AAY38386-Y38498)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 36
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AA206219 for described uses).
XX
SQ Sequence 27 AA:
XX
OY Query Match 3.5%; Score 127; DB 20; Length 27;
Best Local Similarity 88.9%; Pred. No. 0.00023;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 76 RLESKYGSLYRRNVILSATHTSGPA 102
1 RLOSKYGSLYRRNVILSGTHTSGPA 27
XX
RESULT 14
ID AAY38466 standard; Protein: 26 AA.
XX
AC AAY38466;
XX
DT 18-OCT-1999 (first entry)
XX
DE Human secreted protein encoded by gene No. 21.
XX
KM Human; secreted protein; fusion protein; gene therapy; protein therapy;
KM diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KM developmental abnormality; foetal deficiency; blood; allergy; renal;

KM immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KM inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid digestion;
KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO935158-A1.
XX
PD 15-JUL-1999.
XX
PF 06-JAN-1999; 99WO-US00108.
XX
PR 07-JAN-1998; 98US-0070704.
PR 07-JAN-1998; 98US-0070657.
PR 07-JAN-1998; 98US-0070658.
PR 07-JAN-1998; 98US-0070692.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
XX
DR WPI: 1999-444190/37.
DR N-PSDB: AA206239.
XX
PT New isolated human genes and the secreted polypeptides they encode
XX
PS Disclosure; Page 214; 227pp; English.
XX
SS This sequence represents a secreted human protein encoded by the gene
CC clone detailed in the descriptor line. The gene can be used to generate
CC fusion proteins by linking to the gene to a human immunoglobulin Fc
CC portion (e.g. AA206210) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 36 novel genes and their fragments (nucleic
CC acid sequences: AA206219-206263; amino acid sequences AAY38386-Y38498)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 36
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AA206219 for described uses).
XX
SQ Sequence 26 AA:
XX
OY Query Match 3.5%; Score 125; DB 20; Length 26;
Best Local Similarity 84.6%; Pred. No. 0.00033;
Matches 22; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 10 RADCTGVSDINLMGYGKNGQNGRL 35
1 RADCTGVADINLMGYGKSGQNGGI 26
XX
RESULT 15
ID ABB47751 standard; Protein: 2013 AA.
XX
AC ABB47751;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #455.
XX
KM Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KM vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 12:39:23 ; Search time 25 Seconds

(Without alignments)
1131.474 Million cell updates/sec

Title: US-09-937-521-14

Perfect score: 3582
Sequence: 1 FSGYIGVGRADCTGVSDI.....KPAVILAFGIGISFEVVT 682

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	3.2	1331	CTA2_HUMAN	G9uhc6 homo sapien
2	109	3.0	752	CO2_HUMAN	F06681 homo sapien
3	109	3.0	1953	BIGA_SALTY	F25927 salmonella
4	106.5	3.0	665	FLID_BORBU	G51173 borellia hu
5	106.5	3.0	2628	HAGA_PORGI	G51845 porphyromon
6	106	3.0	1376	CRBH_HUMAN	P82279 homo sapien
7	105.5	2.9	440	FLIT_BACSU	P23445 bacillus su
8	104	2.9	1330	XDH_RAT	P22985 rattus norv
9	103.5	2.9	557	PUR6_DEBOC	P50504 debaromyce
10	103	2.9	451	ARP2_PLAFA	P13824 plasmodium
11	103	2.9	1012	POL5_HBVA	R08364 avian infec
12	102.5	2.9	582	HS60_SCHPO	G09864 schizosacch
13	102.5	2.9	1026	STAU_DROME	P25159 drosophila
14	102.5	2.9	1027	AF10_HUMAN	P55197 homo sapien
15	101.5	2.8	754	METE_CAMJE	G9pnt4 campylobact
16	101.5	2.8	4725	DYHC_DICDI	P43036 dictyosteli
17	101	2.8	1399	RPOC_PSEAE	G9hwc9 pseudomonas
18	100.5	2.8	1715	NX2A_RAT	G03374 rattus norv
19	100	2.8	379	METX_LEPME	P94891 leptospira
20	99.5	2.8	669	GALC_CANFA	P54804 canis famli
21	99.5	2.8	1712	NX2A_HUMAN	G9p282 homo sapien
22	98.5	2.7	836	GYRA_MYCGE	P47280 mycoplasma
23	98	2.7	864	TRKC_RAT	G03351 rattus norv
24	98	2.7	868	PRIT_PORGI	P43158 porphyromon
25	97.5	2.7	376	NIR_ALCPA	P38501 alcaligenes
26	97.5	2.7	882	CADI_HUMAN	P38501 alcaligenes
27	96.5	2.7	510	ENP1_HUMAN	P49961 homo sapien
28	96.5	2.7	1292	RPOC_MYCGE	P47582 mycoplasma
29	96.5	2.7	2051	FAS1_YEAST	P07149 s fatty aci
30	96	2.7	637	DNM3_BROV	G05961 bruceella ov
31	96	2.7	1012	POLS_IBOVC	P15480 avian infec
32	96	2.7	1012	POLS_IBOVS	P22351 avian infec
33	95.5	2.7	3130	DPO2_HUMAN	G06073 homo sapien

34	95	2.7	484	YES1_METUA	O58846 methanococc
35	95	2.7	553	HTU1_TREVO	O97824 thermoplasm
36	94.5	2.6	240	GTOU_MOUSE	O09131 mus musculu
37	94.5	2.6	331	DIV_ECOLI	P15286 escherichia
38	94.5	2.6	518	GLVM_SOLTU	P50433 solanum tub
39	94	2.6	431	GLAI_RHIME	O92946 rhizobium m
40	94	2.6	501	DLDH_PEA	P31023 pisum sativ
41	94	2.6	501	GSPE_AERHY	P31741 aeromonas h
42	94	2.6	593	DCPI_PEA	P51850 pisum sativ
43	94	2.6	706	CIMA_PARP	O45358 paenibacill
44	94	2.6	1447	DDC_HUMAN	P43146 homo sapien
45	94	2.6	2415	PGCA_HUMAN	P16112 homo sapien

ALIGNMENTS

RESULT 1
CTA2_HUMAN STANDARD; PRT; 1331 AA.
ID C9uhc6; G9uhc6; 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Contactin associated protein-like 2 precursor (Cell recognition molecule Caspr2).
GN CNTNAP2 OR CASPR2 OR KIA0868.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=20088303; PubMed=10624965;
RA Poliak S., Gollan L., Martinez R., Custer A., Einheber S.,
SA Salzer J.L., Trimmer J.S., Shrager P., Peles E.;
RT "Caspr2, a new member of the neuroligin superfamily, is localized at the
RT juxtaparanodes of myelinated axons and associates with K+ channels.";
RL Neuron 24:1037-1047(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21250995; PubMed=11352571;
RA Nakabayashi K., Scherer S.W.;
RT "The human contactin-associated protein-like 2 gene (CNTNAP2) spans
RT over 2 Mb of DNA at chromosome 7q35.";
RL Genomics 73:108-112(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:355-364(1998).
CC -I- FUNCTION: MAY PLAY A ROLE IN THE FORMATION OF FUNCTIONAL DISTINCT
CC DOMAINS CRITICAL FOR SALVATORY CONDUCTION OF NERVE IMPULSES IN
CC MYELINATED NERVE FIBERS. SEEMS TO DEMARCAE THE JUXTAPARANODAL
CC REGION OF THE AXO-GLIAL JUNCTION.
CC -I- SUBUNIT: ASSOCIATES WITH KCNA2 (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: Type I membrane protein (potential).
CC -I- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN NERVOUS SYSTEM.
CC -I- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
CC -I- SIMILARITY: CONTAINS 4 LAMININ G-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -I- SIMILARITY: BELONGS TO THE NEUREXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AF193613: AAF25199.1: -
 DR EMBL: AF319045: AAK34932.1: -
 DR EMBL: AF318292: AAK49902.1: -
 DR EMBL: AF318298: AAK49903.1: -
 DR EMBL: AB020675: BAA74891.1: ALT_INIT.
 DR HSSP: P12259: ICZT.
 DR Genew: HGNC:13830: CNTNAP2.
 DR MIM: 604569: -
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000421; FA58_C.
 DR InterPro: IPR001092; HLH_basic.
 DR InterPro: IPR001791; Laminin_G.
 DR InterPro: IPR003585; Neurexin-like.
 DR Pfam: PF00008: EGF: 4.
 DR Pfam: PF00054; Laminin_G: 6.
 DR Pfam: PF00754; F5_F8_type_C: 2.
 DR SMART: SM00294: 4.1m: 1.
 DR SMART: SM00181: EGF: 1.
 DR SMART: SM00001: EGF-like: 1.
 DR SMART: SM00231: FA58C: 1.
 DR SMART: SM00282; Lameg: 4.
 DR PROSITE: PS00022; EGF_1: FALSE_NEG.
 DR PROSITE: PS01186; EGF_2: FALSE_NEG.
 DR PROSITE: PS01285; FA58C_1: 1.
 DR PROSITE: PS01286; FA58C_2: 1.
 DR PROSITE: PS50025; LAM_G_DOMAIN: 4.
 DR GlycoProtein: Cell adhesion: Signal: Transmembrane; Repeat.
 DR SIGNAL: 1
 DR CHAIN: 28 1331
 DR DOMAIN: 28 1262 EXTRACELLULAR (POTENTIAL).
 DR TRANSMEM: 1263 1283 POTENTIAL.
 DR DOMAIN: 1284 1331 CYTOPLASMIC (POTENTIAL).
 DR DOMAIN: 38 178 P5/8 type C.
 DR DOMAIN: 216 368 LAMININ G-LIKE 1.
 DR DOMAIN: 401 552 LAMININ G-LIKE 2.
 DR DOMAIN: 534 591 EGF-LIKE 1.
 DR DOMAIN: 827 963 LAMININ G-LIKE 3.
 DR DOMAIN: 963 1002 EGF-LIKE 2.
 DR DOMAIN: 1055 1214 LAMININ G-LIKE 4.
 DR CARBOHD: 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHD: 346 346 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHD: 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHD: 379 379 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHD: 436 436 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHD: 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHD: 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHD: 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHD: 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHD: 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHD: 1116 1116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHD: 1198 1198 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR SEQUENCE: 1331 AA: 14816 MW; CPB2CB55BEFB99C2 CRC64;

Query Match 3.2%; Score 116; DB 1; Length 1331;
 Best Local Similarity 20.5%; Pred. No. 1.5;
 Matches 90; Conservative 60; Mismatches 149; Indels 140; Gaps 22;

QY 49 PDGSRMAFV-----SVELCMISQRLLEVLKRLKESKGLYRRDNYLSAIHHS 99
 DB 335 PSSSRKKNKGCMEISINYNNGVITDLARKKLE-----PSNNG-----NLSFSCVEPYT 373
 QY 100 GAGGFEOTLYT-----LASGFSNRT-----FOYVSGIMKSIDIAHTNKP 142
 DB 374 VPV-FFNATSYLEVGRNLQDLFSVSFOFTNNPGLVFSHPADN-IGNVEIDLTESKV 431
 QY 143 GRIPIKNGVNAVQINRSPSYLLPQSERARYSSNTDQEMVLKLVLDNGEDGLGIS-- 200
 DB 432 G-VHININQTKNSQIDISSGSLNDGOMHEVRFLA---KENAIIITID--GDEASAVRIN 485

QY 201 -----WFAIHVPVSMNNSH-----FVNSDNMGAYLFEQKNKGYL 237
 DB 486 SPLQVTKGKYEFGFGLNMNNSHSLVLPSPFGCMQLQYVDQVNIYEVAQR----- 540
 QY 238 PGSGPFVAGFASNSLGDVSPNLTGPHCVYTG---ESCDNDKSTCPNGSGSMKAGSPGD 294
 DB 541 PGSFANVSTDCATIIDRCVFN---HCEHGKCSQTMDSFKCTCETGIS-----GAT 589
 QY 295 MESTHIIGRIIYOKAKELVYASASQ-----EVTGPFVLAHQMVNNTDVSVOLNATH 345
 DB 590 CHNS-----IYPSCEAYKHLGQTSNYWIDPPGSGPLGLKYCNMTEDKVMYIYSH 642
 QY 346 TVKTCFPAIGYS-----FACTIDVSGSLNITQ-----GTFBCD 379
 DB 643 DLQMQTPVGVNPEKYSVTQLYVSASMDISAITDSAECEQVYSFCKMSRLMTPPGS 702
 QY 380 PF--MDTLRDQLLGRKPEE 396
 DB 703 PYTWM-----VGKANEX 714

RESULT 2
 CO2_HUMAN
 ID CO2_HUMAN STANDARD: PRT: 752 AA.
 AC P06681; Q13904;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Complement C2 precursor (EC 3.4.21.43) (C3/C5 convertase).
 GN C2.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=87127920; PubMed=2949737;
 RA Bentley D.R.;
 RT "Primary structure of human complement component C2. Homology to two unrelated protein families.";
 RL Biochem. J. 239:339-345(1986).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93315633; PubMed=8326124;
 RA Ishii Y., Zhu Z.B., Macon K.J., Volanakis J.E.;
 RT "Structure of the human C2 gene.";
 RL J. Immunol. 151:170-174(1993).
 RN (3)
 RP SEQUENCE FROM N.A.
 RA Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E., Banta A., Swartzell S., Smith T.M., Spies T., Hood L.;
 RT "Sequence determination of 300 kilobases of the human class III MHC locus.";
 RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
 RN (4)
 RP SEQUENCE OF 567-752 FROM N.A.
 RX MEDLINE=8414468; PubMed=6199794;
 RA Bentley D.R., Porter R.R.;
 RT "Isolation of cDNA clones for human complement component C2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:1212-1215(1984).
 RN (5)
 RP SEQUENCE OF 694-752 FROM N.A.
 RX MEDLINE=87102880; PubMed=3643061;
 RA Wu L., Morley B.J., Campbell R.D.;
 RT "Cell-specific expression of the human complement protein factor B gene: evidence for the role of two distinct 5'-flanking elements.";
 RL Cell 48:331-342(1987).
 RN (6)
 RP SEQUENCE OF 137-171; 454-466 AND 574-717.
 RX MEDLINE=85038851; PubMed=6149575;
 RA Gagnon J.;
 RT "Structure and activation of complement components C2 and factor B.";

QY 577 IFVGNPKNSAENQ-----HQT---LTVKEDSVADMQIMYNDASWETREYWH 624
DB 427 -----GSKDGERHAFLODTKALHQVFHEHMLDVSKLTDTICGVNMSANASDOERTPMH 481
QY 625 KGILGSLNAT 634
DB 482 VTIKPKSOET 491

RESULT 3
BIGA_SALTY STANDARD; PRT; 1953 AA.
ID BIGA_SALTY P25928; Q9XC03;
AC P25927; P25928; Q9XC03;
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative surface-exposed virulence protein biga precursor.
OS BIGA OR STM3478.
GN *Salmonella typhimurium*.
OC *Salmonella typhimurium*.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC *Salmonella*.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 14028;
RA Stojiljkovic I., Valentine P., Heffron F.;
RT "Salmonella typhimurium rbs homolog."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SCS01412 / ATCC 700720;
RX MEDLINE=21354948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
LT2."
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE OF 1-765 FROM N.A.
RC STRAIN-LT2;
RX MEDLINE=91100301; PubMed=1987123;
RA Wu J.Y., Siegel L.M., Kredich N.M.;
RT "High-level expression of *Escherichia coli* MNDP-sulfite reductase:
requirement for a cloned cyst plasmid to overcome limiting streheme
cofactor."
RL J. Bacteriol. 173:325-333(1991).
CC -1- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
in positions 414 and 732.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL: AF133696; AAD39458.1; -
DR EMBL: AE008859; AAL22340.1; -
DR EMBL: M64606; AAA27042.1; ALT_FRAME.
DR EMBL: M64606; AAA27043.1; ALT_FRAME.
DR PIR: C39200; C39200.
DR PIR: D39200; D39200.
DR StyGene: SG10437; biga.
KM Virulence; Repeat; Signal; Complete proteome.
FT SIGNAL 27
FT CHAIN 1 27
FT 28 1953
FT 101 252
FT 101 103
DOMAIN 101 252
REPEAT 101 103
1 (INCOMPLETE).

FT REPEAT 104 113 2 (INCOMPLETE).
FT REPEAT 114 122 3 (INCOMPLETE).
FT REPEAT 123 133 4.
FT REPEAT 134 144 5.
FT REPEAT 145 155 6.
FT REPEAT 156 166 7.
FT REPEAT 167 177 8.
FT REPEAT 178 188 9.
FT REPEAT 189 199 10.
FT REPEAT 200 210 11.
FT REPEAT 211 221 12.
FT REPEAT 222 232 13.
FT REPEAT 233 243 14.
FT REPEAT 244 252 15 (INCOMPLETE).
FT CONFLICT 207 207 D -> DRGDDVTPPD (IN REF. 1).
FT CONFLICT 514 514 A -> R (IN REF. 3).
FT CONFLICT 1698 1698 D -> N (IN REF. 1).
FT CONFLICT 1795 1798 OYLE -> ITLO (IN REF. 1).
FT CONFLICT 1836 1837 SA -> T (IN REF. 1).
SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;

Query Match 3.0%; Score 109; DB 1; Length 1953;
Best Local Similarity 19.0%; Pred. No. 9.8;
Matches 119; Conservative 86; Mismatches 238; Indels 184; Gaps 26;

QY 3 GYIYGGRADCTGOVS---DINMGYKNGQNRGLTFLRSHAFILADPDGSGNRMAFVS 59
DB 859 GYV-SVIRADNGSEVNSCDILVATYSNSESDEARRATRASGSAVHNKAGCD-ITLIS 916
QY 60 VELCKMISQRRLRLEVK-----RLESKGSLRRDNVILSAIHSHSPAGFFQYTLVI 111
DB 917 DQPPGSGGIEVPLRWYTHTEFYAMMASDGDVNDGAT-----HLOGAGYGVY--- 968
QY 112 LASEGFSNRTFOYIVSGIMKSIDIAHTNLKPKIFINKGVANVQINRSPSYLLNPOSE 171
DB 969 -ASRGKALNCGNIYLLDGLVFTLD-----DENNTISTYWPQSLYLTG----- 1010
QY 172 RARYSSNTDKEMLVKLVDNGEDGLISWFAIHPVSMNNSNHFVNSDNKGYAAYLPEOE 231
DB 1011 -----SGVVASTDADGD-----ATAINGTITVNNAGCGMALNGCYA 1049
QY 232 KNGYL-----PGGPPVAFGASSN-----LGVSPNIIQ 261
DB 1050 INQGVITLTPADDDVYTGADLVGMALNGCVYINDTSGVINDADYGOAFLDSSSYITIN 1109
QY 262 PHCVNTGESCDNDKSTCPNGPS--MCMASGPGQ-DMFESTHILIGRIIYOKARELYASAS 318
DB 1110 NGSINLNGSPMDPTDSHMGCTPTDKIWIQSLPESGDSPTSDTG---FFTACTLANYG 1166
QY 319 QEYTGEPYLAHQV--NMTDVSVOLNATHYKTKPKPALGYSFAGGTIDVSGNLITOGTTE 377
DB 1167 ETLNGDVVNGGWLNEAGASLTVNGT-----VTJNGCANALANYGTLTD 1210
QY 378 GDFP--WDTLRDOLLGKPESEIECQKPKRILHSEGLTTPHPQMDIYDVQIVTGVSLA 435
DB 1211 ADAISTWHSIFNEADSGITD-----LTLNGDVFTYN----- 1243
QY 436 IAAIPBELT-TMSGRFRERAIKKEPALYGNKDMTVYIAGISVNTYHTTYTYEYQAOQRE 494
DB 1244 -----NSDFTGSIAGTSYQGEI-----VNTGDMTVAEDGKSLVSGSFY-FYNEEDATLTIN 1292
QY 495 AASTIYGPHLSAYIQFLRLAKAINTVYANSSGP-----EPPFER 537
DB 1293 SGSAVSGGE--NTIINLTR-----ANDSLTOVSGTITATNGYSAITTVNGSNDPKWIM 1344
QY 538 NLIASLIPNADRAPI-----GKHFGD 559
DB 1345 NTATGVINGINPAPLNLGRCYNGFN 1371

RESULT 4
FLID_BORBU STANDARD; PRT; 665 AA.
ID FLID_BORBU

[illegible]

```

CC - FUNCTION: AGGLUTININATES ERYTHROCYTES.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY C25.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sib.ch).
CC -----
DR EMBL: U41807; AAB17128.1; -
DR InterPro: IPR001769; Peptidase_C25.
DR Pfam: PF01364; Peptidase_C25; 6.
KW Hemagglutinin; virulence; Hydrolase; Thiol protease; signal; Repeat.
FT SIGNAL 1 24
FT CHAIN 25 2628 HEMAGGLUTININ A.
FT DOMAIN 25 539 PEPTIDASE C25-LIKE 1.
FT DOMAIN 540 995 PEPTIDASE C25-LIKE 2.
FT DOMAIN 996 1451 PEPTIDASE C25-LIKE 3.
FT DOMAIN 1452 1907 PEPTIDASE C25-LIKE 4.
FT DOMAIN 2074 2628 PEPTIDASE C25-LIKE 5.
SQ SEQUENCE 2628 AA; 283324 MW; 61C4DE32540C99DA CRC64;

Query Match 3.0%; Score 106.5; DB 1; Length 2628;
Best Local Similarity 18.8%; Pred. No. 24;
Matches 116; Conservative 82; Mismatches 236; Indels 183; Gaps 31;

OY 101 PAGF---FOYLYLILASEGFSNRTFOYIVSGIMK-----SIDIAHTNLKPKIFL--- 147
DB 105 PAGLVDPREYKPVANADASFPTNF--VLDGTAASADIPAGTYDVIINPFIITIVGSG 162
OY 148 -NKGVANVQINRSPSSYLINQSEARARYSNTDKEMLVKLVDINGEDLGILSPALHP 206
DB 163 VSKGNDYVVEAK---TYHFTVQROGP-----GDAASYV----- 193
OY 207 VSMNNSNHFVNSDNGYAAVLFEQKKNKGYLGGPFPVAFASNLGDVSPNIIIGPHCVN 266
DB 194 VTGEGNEFAPVQNLQMSV-----SQGVYLLVQAPASDKRYIVL----- 233
OY 267 TGESCDNDKSTCPNGGSPKCMASGPGDMFESTHTIGRIYOKAKELVASQSEYTPVL 326
DB 224 -NESDFT--QTLPNQM-PMIDADGGHMLSTIN-----YNTAHTQDGC-AM 276
OY 327 AAHQVNMNTDVSVOLNATHYKTKPA-----LQISPAAGTIIDGVS---GLNTOGTE 377
DB 277 FSKSWTASGAKIDISPNIYLVTPKVTVPENKLSY-WVSSQVPTNHHYGVFLSTTQNE 335
OY 378 GDPFMDTLRDQLGKRSSEIVCOKPKPI-LIHSSELTPIHMQODIYDVQVLTGSLAI 436
DB 336 AANFTIKLLEETLGS-----DKPAPNVLKSGGVKLPAFYQERTID-----L 377
OY 437 AAIPELTTMSGRRRREAIKKEFALYGMKMTVTVAGLSNVYTHITTYEEYOQAOYEVA 496
DB 378 SAVAGQOYVLAFRHNST--GIFRLY-LDDVAVSEGGSDNTTY--TYRDMVVAQNLA 432
OY 497 STTY-----GPHLSAVIQLFRDLAIAIATDVANMSSGPEPFKNLALIPNIAD 549
DB 433 ATTFQENAVAPQOYVCVEKXTAGVSPKCVKDVVEGSGN-EFAHVONLITGS----- 483
OY 550 RAPICKHFGDVLQPAKPEYRGVEVVEYLFVGANPRNSAENQHOFLIYVEKEDSV-ADM 608
DB 484 -----AVGQKVTLMW-----DADGNPPNPPGTTTLTSESENGIPMSW 521
OY 609 QIMYND---SWETRFYHMKGILGLSNATYIWHIDTAVPGIYRTYFGHNKQELLAPA 665
DB 522 KTIADADGGNNMTT-----TPPPG--GISFAGHN-SALCASSA 556
OY 666 VILAFEGISSPPEVVT 682
DB 557 SYINFEQPNDPNYLV 573

```

```

RESULT 6
ID CRB1_HUMAN STANDARD; PRT; 1376 AA.
AC P82279;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Crumbs protein homolog 1 precursor.
GN CRB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS RP12 V-161, W-250, M-745; C-764;
RP Y-948; T-1041 AND P-1071.
RC TISSUE=Retina, and fetal brain;
RX MEDLINE=94938399; PubMed=10508521;
RA van den Born L.I., van Driel M.A., de Kok Y.J.M., van Soest S.,
RA Bhatnagar S.S., Kellner U., Hoyng C.B., Westerveld A.,
RA Brunner H.G., Bleeker-Wagemakers E.M., Deutman A.F.,
RA Heckmatt J.R., Cremers F.P.M., Bergen A.A.B.;
RT "Mutations in a human homologue of drosophila crumbs cause retinitis
RT pigmentosa (RP12)".
RT Nat. Genet. 23:217-221(1999).
RL [2]
RP VARIANT LCA ARG-1100, AND VARIANT RP ARG-1181.
RX MEDLINE=21303018; PubMed=11389483;
RA de Kok Y.J.M., van der Velde-Vissers S.D., Kellner U., Jankies B.,
RA van Schooneveld M.J., Blankenagel A., Kohnschneider K., Wissing B.,
RA Cruijsberg J.R.M., Deutman A.F., Brunner H.G., Apfelstedt-Sylla E.,
RA Hoyng C.B., Cremers F.P.M.;
RT "Leber congenital amaurosis and retinitis pigmentosa with Coats-like
RT exudative vasculopathy are associated with mutations in the crumbs
RT homologue 1 (CRB1) gene."
RL Am. J. Hum. Genet. 69:198-203(2001).
CC - FUNCTION: MAY BE INVOLVED IN CELL-CELL INTERACTION IN NEURONAL
CC DEVELOPMENT OF THE RETINA.
CC - SUBCELLULAR LOCATION: Extracellular (Potential).
CC - TISSUE SPECIFICITY: PREFERENTIAL EXPRESSION IN RETINA, ALSO
CC EXPRESSED IN BRAIN AND FETAL BRAIN.
CC - DISEASE: DEFECTS IN CRB1 ARE THE CAUSE OF RETINITIS PIGMENTOSA
CC TYPE 12 (RP12), AN AUTOSOMAL RECESSIVE CONDITION WHICH IS
CC CHARACTERIZED BY NIGHT BLINDNESS FROM EARLY CHILDHOOD AND
CC PROGRESSIVE VISUAL FIELD LOSS. THERE IS GENERAL LOSS OF RETINAL
CC PIGMENT EPITHELIUM THROUGHOUT THE RETINA AND PATIENTS EXPERIENCE
CC SEVERE VISUAL IMPAIRMENT BEFORE THE AGE OF TWENTY.
CC - DISEASE: Defects in CRB1 are a cause of a form of Leber congenital
CC amaurosis (LCA).
CC - SIMILARITY: CONTAINS 15 EGF-LIKE DOMAINS.
CC - SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC - DATABASE: NAME-Mutations of the CRB1 gene;
CC NOTE-Retina International's Scientific Newsletter;
CC WWW="http://www.retina-international.com/sci-news/crb1mut.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sib.ch).
CC -----
DR EMBL: AF154671; AAF01361.1; -
DR HSSP: P08709; 1BP9.
DR Genew; HGNC:2343; CRB1.
DR MIM: 604210; -
DR InterPro: IPR000152; Asx_hydroxy1.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.

```


CC 287 MASGPEQDMFESTHITIGRIITVOKAKELYASASOETGPVLAHQWNNMDVSVOLNATHT 346
DB 901 -----VCHSRMDPFCSCSPALITGKACEEYQWCMGCFSP-----CPHG 936
QY 347 VKTCKPAL-GYSPAGTI-DGVYGL-----NITGGT 376
DB 937 AC-COPVLOGFECIANAVFNGOSGQILFRSNGNITRELT 974

RESULT 7
FLIT_BACSU
ID FLIT_BACSU STANDARD: PRT: 440 AA.
AC P23445;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Flagellum-specific ATP synthase (EC 3.6.3.14).
GN FLIT
OS Bacillus subtilis.
OC Bacteria: Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=91258343; PubMed=1828465;
RA Albertini A.M., Carmori T., Crabb W.D., Scoffone F., Galizzi A.;
RT "The fla locus of Bacillus subtilis is part of a large operon coding
RT for flagellar structures, motility functions, and an ATPase-like
RT polypeptide."
RT J. Bacteriol. 173:3573-3579(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Bourrier L., Brans A., Brun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritsch C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giesepi H., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kuno M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetlelle D., Portwilk S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolet C., Roche E., Roche B., Rose M., Satae Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Toseki V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
RA Viari A., Mamuti R., Medler E., Medler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoko K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danich A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- FUNCTION: PROBABLE CATALYTIC SUBUNIT OF A PROTEIN TRANSLOCASE FOR
CC FLAGELLUM-SPECIFIC EXPORT, OR A PROTON TRANSLOCASE INVOLVED IN
CC LOCAL CIRCUITS AT THE FLAGELLUM.
CC -1- CATALYTIC ACTIVITY: ATP + H(2O) + H(+) (in) -> ADP + phosphate +
CC H(+) (out).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).
CC -----
DR EMBL: X56049; CA39523.1; -
DR EMBL: 299112; CAB13497.1; -
DR PIR: S14497; PMBSAS.
DR PIR: D42365; D42365.
DR Subtilisin; BG10243; flit.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR004100; ATPase_a/bn.
DR InterPro: IPR000194; ATPase_a/bcentre.
DR Pfam: PF00006; ATP-synt_ab_1.
DR Pfam: PF02874; ATP-synt_ab_N_1.
DR SMART: SM00382; AAA; 1.
DR TIGRPFAM: TIGR01026; flit_lysN; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolase; Hydrogen ion transport; ATP synthesis; ATP-binding;
KW Transport; Protein transport; Flagella; Complete proteome.
FT NP_BIND 167 174 YATAAEYFRKGQVVMFPMDSVTFVMAAOREIGLAGEPPTKGYPS
SQ SEQUENCE 440 AA; 47870 MW; 082F2085AF5D8A4D CRC64;
Query Match 2.9%; Score 105.5; DB 1; Length 440;
Best Local Similarity 23.0%; Pred. No. 1.8;
Matches 59; Conservative 40; Mismatches 67; Indels 91; Gaps 15;
QY 357 SEPAQTIDVSG--LNTGGTEGDPFWTLRDQL--GKSEIIVECQKPKPILHSQ 411
DB 80 SLAPGSIVAEATGSLRKVGT-----GLIGQVVDARGEPIDBSF--CRKSPV---ST 127
QY 412 ELTTPHW-QPDL-----VDQVITVG----- 432
DB 128 EOSPPNPKRPPLPREMGVQNSID-SLTVGKGORIGTFPASGCKSTLMGIAKQTEA 186
QY 433 SLTAIAIPSELTTMSGRFREAIKKEFALYGMKDMTVIAGLSNVTYITTYEYQAO 492
DB 187 DLNVVALVE---RREREFEIKDLGKGLRSIVVA-----TSDQPALMR 231
QY 493 YENASTITGPHLSATIQFLDLKAI--ATDVANMS-----SGEPPPEFKL--- 539
DB 232 LKNA-----YTATAAEYFRKGQVVMFPMDSVTFVMAAOREIGLAGEPPTKGYPS 285
QY 540 IASLIPNIDRAPIGKH 556
DB 286 VFAILPRLRLERGANEH 302

RESULT 8
XDH_RAT
ID XDH_RAT STANDARD: PRT: 1330 AA.
AC P22965; O63157;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Xanthine dehydrogenase/oxidase [includes: Xanthine dehydrogenase
DE (EC 1.1.1.204) (XO); xanthine oxidase (EC 1.1.3.22) (XO) (Xanthine
DE oxidoreductase)].
GN XDH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=90354396; PubMed=2387845;
RA Amaya Y., Yamazaki K.-I., Sato M., Noda K., Nishino T., Nishino T.;
RT "Proteolytic conversion of xanthine dehydrogenase from the

DR InterPro: IPR002663; Birna_VP3.
 DR InterPro: IPR002664; Birna_VP4.
 DR Pfam: PF01766; Birna_VP2; 1.
 DR Pfam: PF01767; Birna_VP3; 1.
 DR Pfam: PF01768; Birna_VP4; 1.
 KW Polypeptide: Structural protein; Nonstructural protein; Hydrolase;
 FT CHAIN 1 452 MAJOR STRUCTURAL PROTEIN VP2.
 FT CHAIN 722 NONSTRUCTURAL PROTEIN VP4 (PROTEASE).
 FT CHAIN 724 1012 MINOR STRUCTURAL PROTEIN VP3.
 SQ SEQUENCE 1012 AA: 109503 MW: D9320A90459DE86 CRC64;
 Query Match 2.9%; Score 103; DB 1; Length 1012;
 Best Local Similarity 20.5%; Pred. No. 10;
 Matches 104; Conservative 70; Mismatches 185; Indels 148; Gaps 26;
 QY 231 ENKNGYLPQGGPRVAFASS-----NLGDVSPNT-LGRCHVTGSGCNDK----- 215
 DB 153 DRIGNLVGEVGVLSLPTSYDGLVRLDPIPAIDPKMATCDSSPRRYTTAAD 212
 QY 276 -----STCPNGGSPMCMASGPGODMEFTHIIRIYOKAKELYASASDEVTPVLAHQ 330
 DB 213 DYQSSQYOPGCTTILFSA-NIDAITNLSVGEIVFQTS-----VQGLVL----- 257
 QY 331 WNMMDVSVQLANHTVTKCKPALGYSFAAGTIDVSGNLTGTTGEGDFPMDLRDOL 390
 DB 258 --NATLYVGFDCT-IVTTRAVAGNGTAGT-DNLMPFLVITPSEIITQPVSTIKLEY 313
 QY 391 -----GKPEELVECKRPKPIILHSELI-----PHWOP-DIVIOIVVGSILAT 437
 DB 314 TSKSGQAGDQD-----SWLASGNLVTHGNYPALRPVTLVNERVATGSAVVV 365
 QY 438 AIPGELTTSGRFRFAIKKEFALY--GKMDTVI-----AGLSNY-THITTYEEY 468
 DB 366 AGVSNFELLNPBELANLTYEGRFPGAMNTIKLLSERDRLGITTWPTREYTFPREY 425
 QY 489 QAQRYEASTIYGPHTLSAYIOLFRLAKAIAITDYANNSGPEPEFKNLIALSPNIA 548
 DB 426 ---FMEVADLNSPLKAGAFG-FKDIIRIRIAPV-----VSTLEP--- 464
 QY 549 DRAPRIKHGVDVLPKRP-EYRGEVVEYF-----VGANPKNSAQNQHOF 595
 DB 465 -----PAAPLAHAGEVDYLLGDGAASGTARAKAASGRIRIOL 510
 QY 596 LVEYKEDSVAD-WQIYNDAWSETRFYHMKGLGTSNATYHIDPTAGYIRIRYRG 654
 DB 511 LAADKGEVYANLFQVPQNPV-----DGLT-----ASPGVLR---GA 545
 QY 655 HN-----RKQELLKPAVILAFEGISSP 676
 DB 546 HNLDCVLRGATLFPVITTVEDAMP 572
 RESULT 12
 HS60_SCHPO STANDARD; PRT; 582 AA.
 ID HS60_SCHPO 009864; 010285;
 AC 009864; 010285;
 DT 01-FEB-1996 (Rel. 33, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Heat shock protein 60, mitochondrial precursor (HSP60).
 GN HSP60 OR MCP60 OR SPAC12612.04.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=HM123;
 RX MEDLINE=96144268; PubMed=8566770;
 RA Yoshida H., Yanaagi H., Yura T.;
 RT Cloning and characterization of the mitochondrial HSP60-encoding

RT gene of Schizosaccharomyces pombe.";
 RL Gene 167.163-166(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=912;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris P., Hildalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Oliver P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Palmer K., O'Neill S., Pearson D., Quail M.A., Rabbittowisch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellern J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voickert G., Aert R., Robben J., Grymonprez B.,
 RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
 RA Beyer P., Zimmermann W., Wedler H., Wandut R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Garriti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 CC - FUNCTION: MAY PARTICIPATE IN ASSEMBLY AND/OR DISASSEMBLY OF
 CC PROTEINS IMPORTED INTO THE MITOCHONDRION. HSP60 ARE ATPASIS AND
 CC HAVE AFFINITY FOR UNFOLDED PROTEINS.
 CC - SUBCELLULAR LOCATION: Mitochondrial.
 CC - SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
 CC - THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z66568; CA91499.1; -;
 DR EMBL: D50609; BAA09171.1; -;
 DR HSSP: P06139; IGRU.
 DR InterPro: IPR001844; Chaprin Cpn60.
 DR InterPro: IPR002423; Cpn60/TCP-1.
 DR Pfam: PF00118; cpn60_TCP1.1.
 DR PRINTS: PR00298; CHAPERONIN60.
 DR PROSITE: PS00296; CHAPERONINS_CPN60; FALSE_NEG.
 KW Chapterone; ATP-binding; Mitochondrion; Transit peptide; Heat shock.
 FT TRANSIT 1 35 MITOCHONDRION (POTENTIAL).
 FT CHAIN 36 582 HEAT SHOCK PROTEIN 60.
 FT CONFLICT 392 392 Y -> S (IN REF. 1).
 FT CONFLICT 415 415 G -> A (IN REF. 1).
 SQ SEQUENCE 582 AA: 62167 MW: E5434436A9BAFC05 CRC64;
 Query Match 2.9%; Score 102.5; DB 1; Length 582;
 Best Local Similarity 23.4%; Pred. No. 4.7;
 Matches 62; Conservative 42; Mismatches 94; Indels 67; Gaps 14;
 QY 133 IDIAHTNLRKPKIFINKGVANYQINRSPSYL-----LNPQSERAR-----YSSN 178
 DB 336 IDVSIKKAQPHH-----GSGGSVTVTKKEPTIIKAGADHKVNDRCQGVADPNLRE 391
 QY 179 TDKEMLVKIVDNGEDGLISWFAIHPMSM-NSNHPVNSDMGTAATLFEDEKRGVLT 237
 DB 392 YEKEKLOERLAKLSG-GIAVIRKGSSEVEVNEKKRIVDALNAVAAY-----SEGVLT 444

QY 238 PGCG-PEVAGFASNNLGDVSPN-----ILGPHCVNTGESCONDKSTCPNGSPSCMASGPG 292
DB 445 PGAGTSTFVK--ASTLRGLDPTNNFDOKLGEIVRA-----LTPRA 483
QY 293 QDMEESTHITGRIYOKAKELYASASQEVTPVLAHQWVMTDVSQVLAHTHYKCKRP 352
DB 484 QTLLENAGLGNLITVGLKELKELKERN--IGYDIADKRFVLDNEIGV-LDPLKVVRT--- 537
QY 353 ALGVSFAGTIDGVSGINTQTTE 377
DB 538 -----GLVD-ASGVASLMTTE 553
RESULT 13
STAU_DROME STANDARD: PRT; 1026 AA.
AC P25159;
DT 01-MAY-1992 (Rel. 22, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Maternal effect protein stauften.
GN STAU.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91300552; PubMed=1712672;
RA St Johnston D., Beuchie D., Nusslein-Volhard C.,
RT "Staufen, a gene required to localize maternal RNAs in the Drosophila
RT egg";
RL Cell 66:51-63(1991).
RN [2]
RP STRUCTURE BY NMR OF 579-646.
RX MEDLINE=95354674; PubMed=7628456;
RA Bricot M., Grunert S., Murzin A.G., Proctor M., St Johnston D.;
RT "NMR solution structure of a dsRNA binding domain from Drosophila
RT staufen protein reveals homology to the N-terminal domain of
RT ribosomal protein S5";
RL EMBO J. 14:3563-3571(1995).
RN [3]
RP ERRATUM.
RA Bricot M., Grunert S., Murzin A.G., Proctor M., St Johnston D.;
RL EMBO J. 14:4385-4385(1995).
RN [4]
RP CHARACTERIZATION OF DRBM DOMAINS.
RX MEDLINE=20183617; PubMed=10716936;
RA Micklem D.R., Adams J., Grunert S., St Johnston D.;
RT "Distinct roles of two conserved Staufen domains in oskar mRNA
RT localization and translation";
RL EMBO J. 19:1366-1377(2000).
CC -1- FUNCTION: REQUIRED BOTH FOR THE LOCALIZATION OF MATERNAL
CC DETERMINANTS TO THE POSTERIOR POLE OF THE DROSOPHILA EGG AND
CC FOR BICOID RNA TO LOCALIZE CORRECTLY TO THE ANTERIOR POLE.
CC OSKAR PROTEIN IS REQUIRED TO KEEP OSKAR RNA AND STAUFEN PROTEIN AT
CC THE POSTERIOR POLE.
CC -1- TISSUE SPECIFICITY: POLAR GRANULES AT THE POSTERIOR POLE OF THE
CC OOCYTE, AND BY THE TIME THE EGG IS LAID, AT THE ANTERIOR POLE.
CC -1- DOMAIN: CONTAINS A PROLINE-RICH DOMAIN, THE INSERTION OF THIS
CC DOMAIN IN THE DRBM 2 DOMAIN IS REQUIRED FOR STAU-OSKAR MRNA
CC LOCALIZATION.
CC -1- SIMILARITY: CONTAINS 5 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M69111; AAA73062.1; -
DR PIR; A40315; A40315.
DR PDB; 1STU; 31-JUL-95.
DR FlyBase; FBgn0003520; stau.
DR InterPro; IPR001159; DS_RBD.
DR Pfam; PF00035; dsrm; 5.
DR SMART; SM00358; DSRM; 4.
DR PROSITE; PS50137; DS_RBD; 5.
KW Developmental protein; RNA-binding; Repeat; 3D-structure.
FT DOMAIN 311 378
FT DOMAIN 398 557
FT DOMAIN 578 645
FT DOMAIN 711 781
FT DOMAIN 951 1018
FT DOMAIN 3 77
FT DOMAIN 42 46
FT DOMAIN 47 51
FT DOMAIN 71 77
FT DOMAIN 451 455
SQ SEQUENCE 1026 AA; 110347 MW; AE5B97624BBF7D0B CRC64;

Query Match 2.9%; Score 102.5; DB 1; Length 1026;

Best Local Similarity 18.4%; Pred. No. 11; Matches 135; Conservative 100; Mismatches 255; Indels 245; Gaps 36;

QY 5 YIGVGRADCTGOVSDINLMGYGRKNGQANGLTLRLSPRADILADPGSNMAFVSELCM 64
DB 283 HVTGNIDATGALSNEDTSSSGRGKD-----KTPMCL 315
QY 65 ISQRLREVLKRLSKSGSLYRRDNVLSAIHSGPAGFOYTYTL-----ASGFS 118
DB 316 VNEIARY-----NKTTQHYR-----LTERGPAHCKTQTVTLMDDEYSADGFK 360
QY 119 NRTFOYIVSGIMKSIDIAHTNLKPGKIFINKGVANVQINRSPSYLLNPQSRARYSN 178
DB 361 IKKAQHILAS--KAIEETMKHPPK-----IRSEEG--GPM--RTHTTPT 401
QY 179 TDKEMLVKLVDLNGEDGLISWFAIHVPMNSN-----HFVNSDNNGY----- 223
DB 402 VEINLALMK-----LQRTFYLLDPTQIIPPLDSIVPPEFAGHILTPAGCMQPPP 453
QY 224 -AAYLVFQEKNGKYL-----GGCGPFVAGFASNNLGDVSPNIIIGPHCVNTG 268
DB 454 PRVLAIRQLGNGVPIPSQPMHHPFHGGQRPFPKPSR--FALPPLGAH-VHGG 509
QY 269 ESCDNDKSTCPNGSPSCMASGP-----GDMFESTHITGRIYOKAKELYA----- 315
DB 510 -----PN-GPPSPVTPPSKITLFGKQKFGV--IGRTIQAKKIDAAARLQVL 555
QY 316 -----SASQEV-----TGPVLAHQWVMTDVSQVLAHTHYKCKR---PALGYS 357
DB 556 KTOAISASEEALIEDSMDEGDKKSPISQVH-----EIGIKRNMTVHKKVIREGPAIMKN 609
QY 358 FAAGTIDVSGMLNTOSTTEGDPFMDLROLLGKPSSE--IYECCKPKILLHSGTLTI 415
DB 610 FITACI--VGSIVTBEENG-----KVSKKRAEEMLVLEQLPLPTPK----- 653
QY 416 PHPWOPDIVOVQIVTVGSIAIAIPGL-----TWSGRFRFEAIKKEPALGMDMTV 469
DB 654 -----QPLKRIKTKTPKSAARREGSVSGTGTGTGTGKRPERRK-----RLMPKDKLI 704
QY 470 VIAGLSNVYTHYI--TTYEYO-----AQRYE-----AASTYIGPHT-- 504
DB 705 DMDADADPIKLILOOTTRKEKEPIELIAKNGNETARREFMEVASASGSTARGCNSK 764
QY 505 -----LSAYIOLFDDLKAITDVIYVANS-----SGPEPPFPKNILASILPINDR 550
DB 765 KLAIRNAQAQLFELLEAVOVTPTNETOSSSECTSATMSAVTAPAVEATAREGVAPVA-- 822
QY 551 APIGKHFGDVL-----QPAKPEYRAGVEVIVFVAGNPKASANDQHTQTLVTEKEDSV 605
DB 823 TPVGPMGILLILNQNKPKAKR-----DQIVIVKSVESKEEAKKEVAVAAEENSNN 876

QY 606 ADMQIMVDSMETR 620
DB 877 ANSGDSSNSSGDSQ 891

RESULT 14
AF10_HUMAN STANDARD: PRT; 1027 AA.
ID AF10_HUMAN
AC P55197;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE AF-10 protein.
GN MLLT10 OR AF10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95195207; PubMed=788665;
RA Chaplin T., Ayton P., Bernard O.A., Saha V., della Valle V.,
RA Hillion J., Gregorini A., Lillington D., Berger R., Young B.D.;
RT "A novel class of zinc finger/leucine zipper genes identified from
RT the molecular cloning of the t(10;11) translocation in acute
RT leukemia";
RL Blood 85:1435-1441(1995).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- TISSUE SPECIFICITY: ABUNDANTLY IN TESTIS.
CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL
CC TRANSLOCATION T(10;11)(P12;Q23) THAT INVOLVES MLLT10 AND MLL/HRX.
CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.
CC -1- SIMILARITY: HIGH TO AF17.
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -1- DATABASE: NAME-Atlas Genet. Cytoenet. Oncol. Haematol.;
CC WWW="http://www.intoblogen.fr/services/chromcancer/Genes/AF10.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U13948; AAA79972.1;
DR GeneW; HGNC:16063; MLLT10.
DR MIM; 602409;
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SMO0249; PHD; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 2.
KW Nuclear protein; Zinc finger; Repeat; Chromosomal translocation;
KW Proto-oncogene.
FT ZN_FING 22 74 PHD-TYPE 1.
FT ZN_FING 68 85 C4-TYPE.
FT ZN_FING 134 197 PHD-TYPE 2.
FT DOMAIN 229 240 GLU/LYS-RICH.
FT DOMAIN 766 794 LEUCINE-ZIPPER.
FT DOMAIN 856 891 POLY-SER.
FT SITE 266 266 MLL FUSION POINT (IN ACUTE MYELOID
FT LEUKEMIA PATIENT B).
FT SITE 643 643 MLL FUSION POINT (IN ACUTE MYELOID
FT LEUKEMIA PATIENT C).
FT SITE 680 680 MLL FUSION POINT (IN ACUTE MYELOID
FT LEUKEMIA PATIENT A).
SQ SEQUENCE 1027 AA; 109026 MW; 7C7C0059DD46589 CRC64;
Query Match 2.9%; Score 102.5; DB 1; Length 1027;
Best Local Similarity 19.3%; Pred. No. 11;
Matches 99; Conservative 63; Mismatches 189; Indels 163; Gaps 20;

QY 27 KNGQNRGLTTRLFRAFLADPDGSSNRMAFVSLCMI SORLRLEVKRLKLE-----S 79
DB 485 KNGENVSHLSVASASPTSSVASAAGS-----ITSSSLQSPILLRNGSLQSLVSGSSPVGS 540
QY 80 KYGSLYRDNV-----ILSAITHSGPAGFFOYLYILASEFGSNRTQYIVSGIMK 131
DB 541 EISMQYRHDGACPTTFSELLNLAHNDGSD-----SSTLTQKELKFI--GIYN 586
QY 132 STDIAHTMLKPGKIRINNGVANNVQINRSPSYLLNPOSERARYS-----SNPK 181
DB 587 SMDVA-----VSFPNVVSGSGSTPVSSSHLPQOSSGHLOQVCAISPASVSNAP 636
QY 182 EMLVLEKLVNAGEDLGLISWFAIHPVSNMNSNHEFVSNMGYAAVLPQENKKGYPGQG 241
DB 637 AVATIQAMTLSSSL-----SQAPSIMYCNRRNSSMAALIAQSENNO----- 678
QY 242 PIVAGFASNNIGDVSNNILGPHCVNTGSGCDNDKSTCPNG--GPSMCAS-----GPGQ 293
DB 679 -----TPQDLGDNRNRLVG-----RGSSPRGSLSPRSVPSSLIQIRYDQGN 719
QY 294 DMFESTHIIGRITVYKAKELVASASQ-----EVTGPVIAAHQW-----VN 333
DB 720 SLENLPPVPAASIEQLERQWSEGOQLLEGQTPSDILGMLKSLHQLOVENRRLKEQIKN 779
QY 334 MTD-----VSQNLATHVTCKPALG---YSFAAGTIDVSGSLNTGTEGDPFMD 383
DB 780 LTAKEERQLNLAQSLVFPPTITANPSPSHQHTHTSADTAPTTDLNSKSHIGNSF-- 837
QY 384 TLRDQLLKPSSEIVECKPKPILHSGELTIPHP-----WQPDIVQIVTVGS 433
DB 838 -LPDNLSPVLNODLTSQGSTS-----SSSALSTPPAGQSPAQOGSGVGVQGVNVTGA 893
QY 434 LA-----IAIRGELTMSG 448
DB 894 LASGMQPYTSTIPAVSAVGIIIGLPGNQLAING 927

RESULT 15
METE_CAMJ2E STANDARD: PRT; 754 AA.
ID METE_CAMJ2E
AC O9PN94;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 5-methyltetrahydropteroyltyriglutamate-homocysteine methyltransferase
DE (EC 2.1.1.14) (Methionine synthase, vitamin B12 independent isozyme)
DE (Cobalamin-independent methionine synthase).
GN METE OR CJ1201.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kelley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feilwell T., Holtroyd S.,
RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences";
RL Nature 403:665-668(2000).
CC -1- FUNCTION: CATALYZES THE TRANSFER OF A METHYL GROUP FROM 5-
CC METHYLTETRAHYDROFOLATE TO HOMOCYSTEINE RESULTING IN METHIONINE
CC FORMATION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 5-methyltetrahydropteroyltyl-L-glutamate + L-
CC homocysteine = tetrahydropteroyltyl-L-glutamate + L-methionine.
CC -1- COFACTOR: ZINC. BINDS ONE ION PER SUBUNIT (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.
CC -1- SIMILARITY: BELONGS TO THE VITAMIN-B12 INDEPENDENT METHIONINE
CC SYNTHASE FAMILY.

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL139077; CAB73455.1;
DR InterPro: IPR002629; Methionine_synth.
DR Pfam: PF01717; Methionine_synth.1
KW Transferase; Methyltransferase; Methionine biosynthesis; Zinc; Repeat;
KW Complete proteome.
FT METAL 641 641 ZINC (BY SIMILARITY).
FT METAL 643 643 ZINC (BY SIMILARITY).
FT METAL 726 726 ZINC (BY SIMILARITY).
SQ SEQUENCE 754 AA; 86922 MW; 113AB568E8EFC0D CRC64;

Query Match 2.8%; Score 101.5; DB 1; Length 754;
Best Local Similarity 19.7%; Pred. No. 8.4;
Matches 116; Conservative 84; Mismatches 203; Indels 187; Gaps 31;

QY 174 RSSNTDKEMLVKLVLDNGEDLGISWAIHPVSN--NSNHFVNSDMNGYAALFE-- 229
DB 23 KTFKQSSKEELK---SAKDLRIHWQEIQAGIDFIPSNDESLYDNVLAVALFNIV 78
QY 230 --QEKN-----KGYLPQGFVAGFVASSNLGDVSPNIIQPHCVMTGESCDND 274
DB 79 HRTKYNLNLDALDEYFAQSRGY-QGENGCVTALAMKKMPTNTHYLVPECDNAD----- 131
QY 275 KSTCPNGSPGCMASGPGQDMPESTHIIIGRIIYQAKELYASASQEVTSPIVLAHQVMN 334
DB 132 -----IALTGDKIFKE-----YLEAKELGIESKPVLIQ-IFTLFKLIAF 170
QY 335 TDVSVQNLNTHVTKCKPALGYSFAAGTIDVSGNLITQGTGEGDPF--WDTLRDO--LL 390
DB 171 KDKETQ-----KLAKERLNNAYLE-LPDKLNEIKVTWLELD-EPYLVYDLSKEDIALF 221
QY 391 GKPSSEIVECQKPKPILHS--GELTIHPMPQDIYVOIVYGSIAIAIPGELTMSG 448
DB 222 EEFYQELLNHKDKILLOSIFGDL--RDIYPKLIESKFDALG-----LDFIEG 268
QY 449 RRFREAIKKEFALYGMKDMTVVIAGISNVTYHTYITEE-----YQQRV-----EAA 496
DB 269 KOSLALWQO---YGFAPKDKILFGGLNGKNIYANDYAKSLKLIKELQKTONIILNTSC 324
QY 497 STIYGPHTL-----SAVIOLE-----ROLAKAIAIDTVANMSSGPEPPFKNL 539
DB 325 SLHVPYSTEFESKLDSTYLLKLPAPAKERLQELKDL-KEILNSSEENPLFRANQELFKNI 383
QY 540 IASL-----IPNIADRAPIGKHEGVDLQPAKPEYRVGEVEVIFVGA 581
DB 384 PERLDEKVARLALKLKEDFTRTPSFKERALLQKER--LKLPLPTTTIGSPQSDVRS 441
QY 582 N-----PKNSAENQTHQTLVEKEYEDSVADMQIMYNDASWETRFYWHKGIIGLSNAT 635
DB 442 NRLAFKQEKISQNYTE--FNQOKIKECI--QIOEE-----IGL----- 476
QY 636 VWHIDTAVPGIYR-----IRYFGHNKQEL-----LKPAYI 667
DB 477 -----DVLVHGFEFRNDVYFEGENLKGLFTQNGWVQSYGTRCVKRPVI 521
```

Search completed: July 3, 2003, 12:43:42
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 3, 2003, 12:42:40 ; Search time 90 Seconds
(without alignments)
1561.379 Million cell updates/sec

Title: US-09-937-521-14
Perfect score: 3582
Sequence: 1 FSGYIGVGRADCTGVSDI.....KPAVILAFEGISSPEFVVT 682

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: SPREMBL_21:*
 - 2: sp_archaea:*
 - 3: sp_bacteria:*
 - 4: sp_fungi:*
 - 5: sp_human:*
 - 6: sp_invertebrate:*
 - 7: sp_mammal:*
 - 8: sp_mhc:*
 - 9: sp_organelle:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvtruss:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query	Match	Length	ID	Description
1	3582	100.0	756	11	Q9JHE3	Q9JHE3 mus musculus
2	3416	95.4	761	11	Q91XT9	Q91XT9 ratius norv
3	3044	85.0	761	4	Q9NR71	Q9NR71 homo sapien
4	2634	73.5	496	11	Q8R236	Q8R236 mus musculu
5	1509.5	42.1	704	5	Q9VAX0	Q9VAX0 drosophila
6	1416	39.5	818	10	Q9AX93	Q9AX93 oryza sativ
7	1398	39.0	808	10	Q9LNV7	Q9LNV7 arabidopsis
8	1373	38.3	702	5	Q15913	Q15913 dictyostel
9	1249.5	34.9	715	10	Q9SH86	Q9SH86 arabidopsis
10	1123	34.1	705	10	Q9FTL4	Q9FTL4 arabidopsis
11	1120.5	31.3	841	3	Q8XOX9	Q8XOX9 neurospora
12	1095.5	30.6	670	16	Q91596	Q91596 pseudomonas
13	1090.5	30.4	670	2	Q9RH00	Q9RH00 pseudomonas
14	884.5	24.7	637	16	Q06769	Q06769 mycobacteri
15	626	17.5	314	10	Q93Z16	Q93Z16 arabidopsis
16	152	4.2	1794	16	Q8RHH1	Q8RHH1 fusobacteri

17	136	3.8	2742	16	Q8XX21	Q8XX21 ralsstonia s
18	131.5	3.7	1565	2	Q85156	Q85156 photorhabdu
19	131.5	3.7	2013	16	Q92EK2	Q92EK2 listeria in
20	131	3.7	616	16	Q8ZEK3	Q8ZEK3 yersinia pe
21	131	3.7	625	16	Q8R9T6	Q8R9T6 thermoaer
22	118.5	3.3	2013	16	Q8Y9T8	Q8Y9T8 listeria mo
23	116.5	3.3	1417	16	Q9CK92	Q9CK92 pasteurilla
24	115	3.2	732	10	Q9FTG2	Q9FTG2 arabidopsis
25	113.5	3.2	1175	2	Q47955	Q47955 haemophilus
26	113	3.2	2015	16	Q9URD2	Q9URD2 neisseria m
27	112	3.1	3972	16	P73139	P73139 synechocyst
28	111.5	3.1	752	6	Q8S074	Q8S074 pan troglod
29	111.5	3.1	847	15	Q8URQ3	Q8URQ3 human immun
30	110.5	3.1	1167	11	Q88341	Q88341 rattus norv
31	110	3.1	1217	3	P97685	P97685 rattus norv
32	110	3.1	1340	3	Q8TFK0	Q8TFK0 debrayomyce
33	110	3.1	2566	17	Q8TSE7	Q8TSE7 methanosaer
34	109	3.0	577	7	Q95IG1	Q95IG1 homo sapien
35	109	3.0	1736	10	Q81899	Q81899 arabidopsis
36	109	3.0	1975	16	Q9K057	Q9K057 neisseria m
37	109	3.0	1995	16	Q9JY23	Q9JY23 neisseria m
38	108	3.0	752	6	Q8S075	Q8S075 pongo pygma
39	108	3.0	854	15	Q74459	Q74459 human immun
40	108	3.0	1053	2	Q30776	Q30776 nostoc punc
41	108	3.0	1576	13	Q91908	Q91908 xenopus lae
42	108	3.0	2271	13	Q91909	Q91909 xenopus lae
43	107.5	3.0	677	10	Q9S2S6	Q9S2S6 arabidopsis
44	107.5	3.0	1193	16	Q97M47	Q97M47 clostridium
45	107.5	3.0	2327	12	Q9QCE4	Q9QCE4 foot-and-mo

ALIGNMENTS

RESULT 1

Q9JHE3 PRELIMINARY: PRT: 756 AA.
ID Q9JHE3
AC Q9JHE3:
DT 01-OCT-2000 (TRENBLREL. 15 Created)
DT 01-OCT-2000 (TRENBLREL. 15, last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, last annotation update)
DE Neutral ceramidase (Neutral ceramidase).
GN ASAH2 OR BCDASE OR LCDASE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20219171; PubMed=10753931;
RA Tani M., Okino N., Mori K., Tanigawa T., Izu H., Ito M.;
RT "Molecular cloning of the full-length cDNA encoding mouse neutral
RT ceramidase - Novel but highly-conserved gene/family of neutral/alkaline
RT ceramidases-".
RL J. Biol. Chem. 275:11229-11234(2000).
DR EMBL: AB037181; BAA84546.1;
DR EMBL: AB037111; BAA94345.1;
DR MGI: MGI:1859310; Asah2.
SQ SEQUENCE 756 AA; 83508 MM; FPD514E51280D4BE CRC64;

QY	1	FSGYIGVGRADCTGVSDINLMGKNGQARGLTFLFSRAFLADPDGSNNMAFVS	60
DB	75	FSGYIGVGRADCTGVSDINLMGKNGQARGLTFLFSRAFLADPDGSNNMAFVS	134
QY	61	ELCMISQRLKLEVKRLSKGSLYRDNVLSAHTHSAGAGFFQITLYLASEGFSNR	120
DB	135	ELCMISQRLKLEVKRLSKGSLYRDNVLSAHTHSAGAGFFQITLYLASEGFSNR	194

```
QY 121 TFOYIVSGIMKSIDIAHTNLKPKGFIINGGNVANOINSPSSYLLNPOSERARISSND 180
DB 195 TFOYIVSGIMKSIDIAHTNLKPKGFIINGGNVANOINSPSSYLLNPOSERARISSND 254
QY 181 KEMLVKLVLDNGEDGLISMFAIHPVSNMNSNHPFNSDMNGYAALFDEKKKGLPGQ 240
DB 255 KEMLVKLVLDNGEDGLISMFAIHPVSNMNSNHPFNSDMNGYAALFDEKKKGLPGQ 314
QY 241 GPFVAGFASNLGDVSPNLLGPHCVNTGSCDNKSTCPNGSPGSMASGPGDMEESTH 300
DB 315 GPFVAGFASNLGDVSPNLLGPHCVNTGSCDNKSTCPNGSPGSMASGPGDMEESTH 374
QY 301 IIGRIYOKAKELIASASOEYVGPVLAHQWVMTVSVQLNATHVTKCPKALGYSFAA 360
DB 375 IIGRIYOKAKELIASASOEYVGPVLAHQWVMTVSVQLNATHVTKCPKALGYSFAA 434
QY 361 GTIDVSGNLITGTEGDFPMDTLRDQLGKPSSEIVECQKPKPILHSGELTIHPMQ 420
DB 435 GTIDVSGNLITGTEGDFPMDTLRDQLGKPSSEIVECQKPKPILHSGELTIHPMQ 494
QY 421 PDIVDQIVTVGSLAIAIPGELITWNGRRFRFAIKKEPALYGMKDMTVIAGLSNVYTH 480
DB 495 PDIVDQIVTVGSLAIAIPGELITWNGRRFRFAIKKEPALYGMKDMTVIAGLSNVYTH 554
QY 481 YTTTVEEYQAREASTTYGPHTSAYIQLPFDLAKATATDVANMSSGPEPPFKNLI 540
DB 555 YTTTVEEYQAREASTTYGPHTSAYIQLPFDLAKATATDVANMSSGPEPPFKNLI 614
QY 541 ASLIPNIADRAPIGKHFGBVLQPAKPEYRGEVEYIFGANGKNSAENQHOTFTLVEK 600
DB 615 ASLIPNIADRAPIGKHFGBVLQPAKPEYRGEVEYIFGANGKNSAENQHOTFTLVEK 674
QY 601 YEDSVADQIMYNDASMETRFYWHKGIILGSNATIYHIDPTAYPGIYRIYFGHNRKOE 660
DB 675 YEDSVADQIMYNDASMETRFYWHKGIILGSNATIYHIDPTAYPGIYRIYFGHNRKOE 734
QY 661 LKPAVILAFEGISSPEFVYTT 682
DB 735 LKPAVILAFEGISSPEFVYTT 756

RESULT 2
Q91XT9 PRELIMINARY: PRT: 761 AA.
AC 091XT9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Ceramidase.
GN PAPKCD.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID-10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;
RX MEDLINE-21336561; PubMed-11328816;
RA Mitsuake S., Tanai M., Okino N., Mori K., Ichinose S., Omori A.,
RA Iida H., Nakamura T., Ito M.;
RT "Purification, Characterization, Molecular Cloning and Subcellular
RT Distribution of Neutral Ceramidase of Rat Kidney.";
RL J. Biol. Chem. 276:26249-26259(2001);
DR EMBL; AB057433; BAB62033.1; -;
SQ SEQUENCE 761 AA; 83487 MW; 68B91BC78AEB6324 CRC64;
```

```
Query Match 95.4%; Score 3416; DB 11; Length 761;
Best Local Similarity 94.1%; Pred. No. 1.5e-262;
Matches 642; Conservative 25; Mismatches 15; Indels 0; Gaps 0;
```

```
QY 1 FSGYIYGVADCTGVSDINLMGYKNGKNGARGLTLRFSAFIADPDGSRMAFVS 60
DB 80 FSGYIYGVADCTGVSDINLMGYKNGKNGARGLTLRFSAFIADPDGSRMAFVS 139
```

```
QY 61 ELCMISORLLEVLKRLKESYKSLYRRDNVILSAIHTHSGPAGFOYTYLILASEGFSNR 120
DB 140 ELCMISORLLEVLKRLKESYKSLYRRDNVILSAIHTHSGPAGFOYTYLILASEGFSNR 199
QY 121 TFOYIVSGIMKSIDIAHTNLKPKGFIINGGNVANOINSPSSYLLNPOSERARISSND 180
DB 200 TFOYIVSGIMKSIDIAHTNLKPKGFIINGGNVANOINSPSSYLLNPOSERARISSND 259
QY 181 KEMLVKLVLDNGEDGLISMFAIHPVSNMNSNHPFNSDMNGYAALFDEKKKGLPGQ 240
DB 260 KEMLVKLVLDNGEDGLISMFAIHPVSNMNSNHPFNSDMNGYAALFDEKKKGLPGQ 319
QY 241 GPFVAGFASNLGDVSPNLLGPHCVNTGSCDNKSTCPNGSPGSMASGPGDMEESTH 300
DB 320 GPFVAGFASNLGDVSPNLLGPHCVNTGSCDNKSTCPNGSPGSMASGPGDMEESTH 379
QY 301 IIGRIYOKAKELIASASOEYVGPVLAHQWVMTVSVQLNATHVTKCPKALGYSFAA 360
DB 380 IIGRIYOKAKELIASASOEYVGPVLAHQWVMTVSVQLNATHVTKCPKALGYSFAA 439
QY 361 GTIDVSGNLITGTEGDFPMDTLRDQLGKPSSEIVECQKPKPILHSGELTIHPMQ 420
DB 440 GTIDVSGNLITGTEGDFPMDTLRDQLGKPSSEIVECQKPKPILHSGELTIHPMQ 499
QY 421 PDIVDQIVTVGSLAIAIPGELITWNGRRFRFAIKKEPALYGMKDMTVIAGLSNVYTH 480
DB 500 PDIVDQIVTVGSLAIAIPGELITWNGRRFRFAIKKEPALYGMKDMTVIAGLSNVYTH 559
QY 481 YTTTVEEYQAREASTTYGPHTSAYIQLPFDLAKATATDVANMSSGPEPPFKNLI 540
DB 560 YTTTVEEYQAREASTTYGPHTSAYIQLPFDLAKATATDVANMSSGPEPPFKNLI 619
QY 541 ASLIPNIADRAPIGKHFGBVLQPAKPEYRGEVEYIFGANGKNSAENQHOTFTLVEK 600
DB 620 ASLIPNIADRAPIGKHFGBVLQPAKPEYRGEVEYIFGANGKNSAENQHOTFTLVEK 679
QY 601 YEDSVADQIMYNDASMETRFYWHKGIILGSNATIYHIDPTAYPGIYRIYFGHNRKOE 660
DB 680 YEDSVADQIMYNDASMETRFYWHKGIILGSNATIYHIDPTAYPGIYRIYFGHNRKOE 739
QY 661 LKPAVILAFEGISSPEFVYTT 682
DB 740 LKPAVILAFEGISSPEFVYTT 761

RESULT 3
Q9NR71 PRELIMINARY: PRT: 761 AA.
AC 09NR71;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Mitochondrial ceramidase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID-9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20347271; PubMed-10781605;
RA El Bawab S., Roddy P., Qian T., Bielawska A., Lemasters J.J.,
RA Hannun Y.A.;
RT "Molecular cloning and characterization of a human mitochondrial
RT ceramidase.";
RL J. Biol. Chem. 275:21508-21513(2000);
DR EMBL; AF250847; AAF86240.1; -;
SQ SEQUENCE 761 AA; 83239 MW; 04F881EDD34A27E4 CRC64;
```

```
Query Match 85.0%; Score 3044; DB 4; Length 761;
Best Local Similarity 82.2%; Pred. No. 5.7e-233;
Matches 560; Conservative 63; Mismatches 58; Indels 0; Gaps 0;
```

```
QY 1 FSGYIGVGRADCTGOVSDINIMMGYKGNQNAKGLTFLRFLRAFLADFDGSDNRMAFVSU 60
Db 80 FSGYHIGGRADCTGOVADINLMGKSGQNAOGLTFLRFLRAFLADFDGSDNRMAFVSU 139
QY 61 ELCHISORLEVLKRLKESKYSGLYRDNVLSAITHSGPAGFPQYTYLASEGFSNR 120
Db 140 DIGVNSORLEVLNRLOSKYSLYRDNVLSGTHSHSGPAGFYQYVFIASGFSNQ 199
QY 121 TFOYIVSGIMKSIDIAHTNLKPGKIFLNKGNVAVNVRSPSYLNLNQSRRARSSMTD 180
Db 200 TFOYIVSGIMKSIDIAHTNLKPGKIFLNKGNVAVNVRSPSYLNLNQSRRARSSMTD 259
QY 181 KEMVLKLVLDNGEDGLISWFAIHPYSMNNSNHFVSDNMGYAAYLFEQEKNGCYLFGQ 240
Db 260 KEMVLKLVLDNGEDGLISWFAIHPYSMNNSNHFVSDNMGYAAYLFEQEKNGCYLFGQ 319
QY 241 GPFVAGRASSNLGVSPNIIAGPCVNTGESCDNKSCTCPNGSPKMASSGPGQDMFESTH 300
Db 320 GPFVAGRASSNLGVSPNIIAGPCVNTGESCDNKSCTCPNGSPKMASSGPGQDMFESTH 379
QY 301 IIGRLIYOKAKELYASASQEVTPVLAHQVNMNTDVSVOLNATHVTCTCPALGYSFAA 360
Db 380 IIGRAMTORAKELYASASQEVTPVLAHQVNMNTDVSVOLNATHVTCTCPALGYSFAA 439
QY 361 GTIDVSGNLITGCTTGGDPFMDLRLQDLGKPSSEIVECQKPRPILHSGELTIPHPWQ 420
Db 440 GTIDVSGNLITGCTTGGDPFMDLRLQDLGKPSSEIVECQKPRPILHSGELTIPHPW 499
QY 421 PDIVDVOIVYGSALAIAPBELITTMGRPREAIKKEFALYGMKDMTVIAGLSNYTH 480
Db 500 PDIVDVOIVYGSALAIAPBELITTMGRPREAIKKEFALYGMKDMTVIAGLSNYTH 559
QY 481 YITTEEXQAOERYEAASITVGPHTLSAYIOLFRLAKAIADTVANSSGPEPPEFKLI 540
Db 560 YITTEEXQAOERYEAASITVGPHTLSAYIOLFRLAKAIADTVANSSGPEPPEFKLI 619
QY 541 ASLIPNIADRAPICKHFGDVLQPAKPEYRGEVEVIFVGNAPKNSAENQTHQFLVEK 600
Db 620 VPLIPSIYDRAPKGRFTGDVLOPAKPEYRGEVEVIFVGNAPKNSAENQTHQFLVEK 679
QY 601 YEDSVADMQIMYNDAWMTREYWHKGLIGLSNATIIYHIPTAPGIRIRIFGHNRKOE 660
Db 680 YEASTSQWICNDASMETREYWHKGLIGLSNATIIYHIPTAPGIRIRIFGHNRKOD 739
QY 661 LLKPAVILAFEGISSPFEVYT 681
Db 740 LLKPAVILAFEGISSPFEVYT 760

RESULT 4
Q8R236 PRELIMINARY: PRT: 496 AA.
AC Q8R236:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to N-acylsphingosine amidohydrolase 2 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER.
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022604; AAH22604.1;
KW Hydrolase.
FT NON TER
SQ SEQUENCE 496 AA: 54554 MW: 0206597694D128C6 CRC64:
Query Match 73.5%; Score 2634; DB 11; Length 496;
Best Local Similarity 100.0%; Pred. No. 1.le-200;
```

```
Matches 496; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 KLVLDNGEDGLISWFAIHPYSMNNSNHFVSDNMGYAAYLFEQEKNGCYLPGQPFVAG 246
Db 1 KLVLDNGEDGLISWFAIHPYSMNNSNHFVSDNMGYAAYLFEQEKNGCYLPGQPFVAG 60
QY 247 FASSNLGVSPNIIAGPCVNTGESCDNKSCTCPNGSPKMASSGPGQDMFESTHIGRII 306
Db 61 FASSNLGVSPNIIAGPCVNTGESCDNKSCTCPNGSPKMASSGPGQDMFESTHIGRII 120
QY 307 YOKAKELYASASQEVTPVLAHQVNMNTDVSVOLNATHVTCTCPALGYSFAAGTIDGV 366
Db 121 YOKAKELYASASQEVTPVLAHQVNMNTDVSVOLNATHVTCTCPALGYSFAAGTIDGV 180
QY 367 SGLNTGCTTGGDPFMDLRLQDLGKPSSEIVECQKPRPILHSGELTIPHPWDIVDV 426
Db 181 SGLNTGCTTGGDPFMDLRLQDLGKPSSEIVECQKPRPILHSGELTIPHPWDIVDV 240
QY 427 QIVTVGSLAIAIAPBELITTMGRPREAIKKEFALYGMKDMTVIAGLSNYTHIITYE 486
Db 241 QIVTVGSLAIAIAPBELITTMGRPREAIKKEFALYGMKDMTVIAGLSNYTHIITYE 300
QY 487 EYQAOERYEAASITVGPHTLSAYIOLFRLAKAIADTVANSSGPEPPEFKNLASLIPN 546
Db 301 EYQAOERYEAASITVGPHTLSAYIOLFRLAKAIADTVANSSGPEPPEFKNLASLIPN 360
QY 547 IADRAPICKHFGDVLQPAKPEYRGEVEVIFVGNAPKNSAENQTHQFLVEKEDSVA 606
Db 361 IADRAPICKHFGDVLQPAKPEYRGEVEVIFVGNAPKNSAENQTHQFLVEKEDSVA 420
QY 607 DMQIMYNDAWMTREYWHKGLIGLSNATIIYHIPTAPGIRIRIFGHNRKOEILLKPAV 666
Db 421 DMQIMYNDAWMTREYWHKGLIGLSNATIIYHIPTAPGIRIRIFGHNRKOEILLKPAV 480
QY 667 ILAFEGISSPFEVYT 682
Db 481 ILAFEGISSPFEVYT 496

RESULT 5
Q9VA70 PRELIMINARY: PRT: 704 AA.
AC Q9VA70:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE CG1471 protein.
DE CG1471.
GN Drosophila melanogaster (Fruit fly).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RA MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaratides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Barton R.C., Rogers Y.-H.C., Blazer R.G., Chumpe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
```

RA Glodok A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Modarity C., Morris J., Moshirel A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo J., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003774; AAF57052.1;
 DR Flybase: FBgn0039774; CG1471.
 SQ SEQUENCE 704 AA; 78231 MW; E65F6IDFLAC2E455 CRC64;

Query Match 42.1%; Score 1509.5; DB 5; Length 704;
 Best Local Similarity 44.6%; Pred. No. 4,8e-111;
 Matches 311; Conservative 114; Mismatches 230; Indels 43; Gaps 14;

QY 4 YIICGRADCTGCVSINIMCGKNGQANRGLTLRFSRAFLIADPDGSRMAFVSVEIC 63
 Db 25 YVVGGRADITGPPEVIMGYANIKQVGRGITHVFAFAFVEDEKG-NRAFAVSADAG 83
 QY 64 MISORLRLVLRLESKYSLRRDNVILSAIHTSGPAGFOYTYILASEGFSRTQ 123
 Db 84 MNGYGLKREVIRKQARVYNTIYNDVAISGTHHAPGCGFLMLHLTSLIGFYQTE 143
 QY 124 YVSGIMKSIDIAHTNLKPGKIFINKNVANQINRSPSSYLLNPOSERARVSSNTDKM 183
 Db 144 VNAQGYLCIKRATNLDVGRILLSTKTVLNINRSPSSYLLNPAEERAOYEHDPKTL 203
 QY 184 LVKLVDLNGEDGLISMPAIFHPVSNNSNHFVNSNMGYAAYLFEOEKNGKGLPGQGF 243
 Db 204 TOLRFDLENNLLGAFNMAVHATSNMNNRNLVTSQNVGYAALLLEKEVNPMPKPKGF 263
 QY 244 VAGFASNNLDVSPNIPHCYNTGESCDNDKSTCPNGSPCMGSPQDMFESHIIIG 303
 Db 264 VAFCCSNIGDVSPTNMGPKCSISGNECDLLTSRCPT-GEQDCAFSGPKDMFESHIIIG 322
 QY 304 RIITYOKA-----KELYASASQEVTPVLAHQVMNMTDVSVOLNATHT-----VKTC 350
 Db 322 QRLADAALGLLNEQSOESTAREVTDVRFIHOVDMPNPNY--GSTYNPLSRKVRKIRGC 379
 QY 351 KALCYSAFAGTIDGVSGLNTGTEGSPFMDTLRDOLLRKPSSEIVECOKPKPILHS 410
 Db 380 QPAMGVSFAAGTTDGGAFSFEQGTTPDMNPNFAD-FIAAPTOEDICHEKPKILLAT 438
 QY 411 GELTIPHPMODIVDQIVTVYVGSIAIAIPGELTITSGRRPREAKFEKALYGMKMTAV 470
 Db 439 GATPEYEMQKPIVSOQLIKIGDIVIAAPCEFTTAGRRRLNQHIAAASAVAGIDTEYI 498
 QY 471 INGLSNVYTHYITTYEEYOARVASTIYGPHTLSAVYQLFRDLAKAIAITDTVANSSG 530
 Db 499 INGLTNYSTYVTEEYOAREEAASITFGPHHSIYMDVFERLTKAMRN--ETVDAG 556
 QY 531 PPPPPKNIASLIPRIA--DRAPIGHREDVLOPAKPEYRGEVNVIVGANPKSAEN 589
 Db 557 PPPPPNMDVLSLNTGVLFDPHPINTDFGYVKSQPNKEKGINEYKVVYISGPRNLT-- 614
 QY 590 QTHQFLVYEEK--YEDSVADMOIMYNDASMETRFYHK--GILGSLNATIIYHIPTAYP 645
 Db 615 FLEKTYETIERKINEDR--WKAVYTDASHMETKMMHKNITILGSEMDIYNDISQTLIP 671

QY 646 GIYRIYFGHNRKQELLKPAVILA-----EGISSPPEV 679
 Db 672 GEXRIHSEYK-----YILGCKYEPYEGJLHSTTV 701

RESULT 6
 Q9AX93 PRELIMINARY; PRT; 818 AA.
 ID Q9AX93
 AC Q9AX93
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE P0501G01.24 protein.
 GN P0501G01.24.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriatloideae; Oryzaceae; Oryza.
 OX NCBI_TaxId=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE.
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0501G01."
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF002819; BAB21095.1;
 SQ SEQUENCE 818 AA; 90267 MW; BFB5436431231449 CRC64;

Query Match 39.5%; Score 1416; DB 10; Length 818;
 Best Local Similarity 39.8%; Pred. No. 1.7e-103;
 Matches 304; Conservative 124; Mismatches 228; Indels 108; Gaps 17;

QY 2 SGYIGVGRADCTGCVSINIMCGKNGQANRGLTLRFSRAFLIADPDGSRMAFVSVE 61
 Db 39 SPFLVGMGSPDITGPAAVNMNGYANTEDJAGIHFRLSRFAFYAEPNG-KRVFVND 97
 QY 62 LCMISORLRLVLRLESKYSLRRDNVILSAIHTSGPAGFOYTYILASEGFSRTQ 121
 Db 98 ACMASQIVIKVLERLKARVGDLYNENNAISGITHHAPGCGYLVVYIVTSIGFVRQS 157
 QY 122 FOYISGIMKSIDIAHTNLKPGKIFINKNVANQINRSPSSYLLNPOSERARVSSNTDK 181
 Db 158 FQYIVDGIQISIVEAHNNLRPKKIFVKNKDLADAGNRSAPAYLNNPAEERKSEYENDK 217
 QY 182 EMVLKLVLDNGEDGLISMPAIFHPVSNNSNHFVNSNMGYAAYLFEOEKNGKGLPGOG 241
 Db 218 EMTLIRFVDELGPVGSFWMFATHGISMSTRNLSISGDKGAARFEMDMAEQMLPKOS 277
 QY 242 ----- 241
 Db 278 AHANSDDLRLSKHTSVLPRAVSTIIPENPEITDILLQLASYSYASGRRLAGSSITRRIR 337
 QY 242 -----PPVAGFASNNLDVSPNIPHCYNTGESCDNDKSTCPNGGSGSMASGPG-Q 293
 Db 338 STQQNKPKFVSAFCCSNCGDVSPNVLGTCIDTNLPCDHNSTG-NGKNELCYGRGPGP 396
 QY 294 DMFESTHILGRIITYOKAKELYASASQEVTPVLAHQVMNMTDVSVOLNAT-----HTVT 349
 Db 397 DEFESTRVIGNQKFLKARLPSASSEIOGKIDIRHTIYIDFSKLEKVKYSTSGCGQYVT 456
 QY 350 CRPALCYSAFAGTIDGVSGLNTGTEGSPFMDTLRDOLLRKPSSEIVECOKPKPILH 409
 Db 457 CPAAMGFAPAAAGTTDGGAFDRDQGVKNPNPKLVNR-LTKTPGQDYECSPKPIILD 515
 QY 410 SGEITIPHPMODIVDQIVTVYVGSIAIAIPBELITITSGRRPREAK-----KEPALX 462
 Db 516 TGEKPEYDMPARLIPVOMIRIGQVLVILCPDEFETMAARRLDVAKTVLTSGNSEFD-- 573
 QY 463 GKMDDTVLAGSNVYTHYITTYEEYOARVASTIYGPHTLSAVYQLFRDLAKAIAITD 522
 Db 574 --KNHIVYLAGLTNSISQIITTFEYQIORYEGASTIYGPHTLSAVYQLFRDLAKAIA 628

QY 523 TVAN--MSSGPEPEFKLILASILPNIA-DRAPIGKHGVDYQ--PAKPEYRGVVEVI 577
 Db 629 -IAKEVETNFPDMDKQGLLPVGFVDSSTPLGKFGVDVNSDPGKSTNKGSTVAT 687
 QY 578 FVGANPKNSAENQHTQFLVEKEDSVADWQIYNDAWETREYWK--GILGSLNATI 635
 Db 688 FYSACPRN--DLTDFGFALVEKL-DGNMNVVYVDDDDMSLRFKMSRPAFLSSRFATL 744
 QY 636 YMHPTDPAVPGIYRIRYFGHNRKQELKP--AVILAEGISSPF 677
 Db 745 EMTVEDAAAGVYRLRHFGAS-----KPMFGSVRHFTGTSRAF 782

RESULT 7
 09LNV7 PRELIMINARY: PRT: 808 AA.
 AC 09LNV7;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE F22G5.28.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
 RA Kim C., Altati H., Bel O., Chin C., Chlou J., Koo T., Lam B., Lee J.,
 RA Conway A., Gonzales A., Hansen N., Howling B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thayer A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC F22G5 from chromosome
 I 1";
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC022464; AAF79556.1; -
 SO SEQUENCE 808 AA; 90243 MW; 6288C800CF1AE04F CRC64;

Query Match 39.0%; Score 1398; DB 10; Length 308;
 Best Local Similarity 39.4%; Pred. No. 4,4e-102;
 Matches 301; Conservative 127; Mismatches 240; Indels 96; Gaps 16;

QY 2 SGYIYGRADCTGOVSDINIMGYGKGNQNRAGLITLFSRAFTLADPGSNRAAFVSVE 61
 Db 36 SEYLIGISYDITGPADVNMAGYANMEQVASGIHFRLRARTFVSEPG-KRVVFNLD 94
 QY 62 LCMISORLRLVLRLESKYSGLRRDNLVLSAHTHSGPAGFOYTYTLASGEFSNRT 121
 Db 95 ACMASQIVKLVIRLAKRGDGLTEOVNIGSIHTHAGPGGYLQYVYIVTSLGFKROS 154
 QY 122 FOYIVSGIMKSIDIAHTNLKPGKIFIKGNVANYQINRSPSSYLLNPGSERARYSSNTD 181
 Db 155 FDALVDGIENSIIDAHNLRGSIFFLNGGELDAGVNRSPSAYILNPKESKSKIRYNDK 214
 QY 182 EMLVLKLVLDNGEDGLISWFAIHPVSMNSNHFVNSDNMGYAALFEQK-EKN----- 233
 Db 215 EMTLLKEVDQMGVGFNMFATHTGTSMTSRTNSLISGDNKGAASRFMEDWDEQNTAERSY 274
 QY 234 -----KGY 236

Db 275 SEETISDEIPRRVSSLIENHODSHLELLASYESOPGKPVTRISSARVRSAALRAD 334
 QY 237 LPGGPVPVAGPSSNLGDVSPNIIIGPHCVNTGESCDNDKSTCPNGGSPMCASGPG--ODM 295
 Db 335 KPG---FVSACQNGCGVSPNVIGARCLDTRGCPDRFNHSTC-GGKNMCMGRGPGYDE 390
 QY 296 FESHTIIGRIYQAKELIYASASOEVTGPVLAHOMVNMIDVSVOL---NATHVTCK 351
 Db 391 FESHTIIGRIYQAKELIYASASOEVTGPVLAHOMVNMIDVSVOL---NATHVTCKPAL 450

QY 352 PALGSPAGATIDVSGSLNITOGTEGDPFMDTLBDOLLGKPSREIVECQPKPILHSG 411
 Db 451 AAMGFAPAGATTDDGADGADFTGGDDKGPFRILYRN-VLKPDPKQIDCHKPKPLLDTG 509
 QY 412 ELTTPHPQPDIVDQIYVGSLSAIAIPGELTMSGRFRFRAIKKEPALYGMKDMT--- 468
 Db 510 EMTKRYDPAKPSILSLQVIRIQGLFSLSPGSETMAGRLRLVAVKTOLKNGNNDLSEI 569
 QY 469 -VVIAGLSNVTHITTYEYQAOQRYEASTIYGPHTLSAVIQLFRLAKAIADPYANM 527
 Db 570 HVVIAGLANGYSOYATPEEYOVORBEASFLYGPHTLSIGIOEFKRLSKSLVLD--MPV 627
 QY 528 SSGPEPP-FFKNLASILPNIADRAPIGKHGVDYQ--PAKPEYRGV--VVEVIFVGANP 583
 Db 628 QPGFQPPDLLKQSLFELPVMMDTPPSDSDGSDVSDVPKLSLRKNGQYTVVFRSACP 687
 QY 584 KNSAENQHTQFLVEKEDSVADWQIYNDAWETREYWK--HNGILGSLNATIYMHLPD 641
 Db 688 RN--DLTDFGFALVEKL-DGNMNVVYVDDDDMSLRFKMSRPAFLSSRFATL 745

QY 642 TAYPGIYRIRYFGHNRK-----QELKPAVILAEGISSPF 679
 Db 746 SASPGVYRIRYFGHNRK-----QELKPAVILAEGISSPF 789

RESULT 8
 015913 PRELIMINARY: PRT: 702 AA.
 AC 015913;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-JAN-1998 (TREMblrel. 05, Last annotation update)
 DE Random slug cDNA25 protein (Fragment).
 GN RSC25.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Iranfar N., Loomis W.F.;
 RA Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U82513; AAB69633.1; -
 FT NON_TER 1 1
 SO SEQUENCE 702 AA; 77000 MW; 213FB11A1A12DA4D CRC64;

Query Match 38.3%; Score 1373; DB 5; Length 702;
 Best Local Similarity 42.0%; Pred. No. 3,4e-100;
 Matches 291; Conservative 120; Mismatches 250; Indels 32; Gaps 17;

QY 2 SGYIYGRADCTGOVSDINIMGYGKGNQNRAGLITLFSRAFTLADPGSNRAAFVSVE 61
 Db 26 SSOYIGIGYDITGPGATNMAGYAMPQITGSIHFRRARAFVFIIDSEG-NRAVYVSTD 84
 QY 62 LCMISORLRLVLRLESKYSGLRRDNLVLSAHTHSGPAGFOYTYTLASGEFSNRT 120
 Db 85 SCMLFQEVKIVIDLOLFEIPPTLYTHONVLSGTHHSGPAGSEIALYITLALGFKK 144
 QY 121 FOYIVSGIMKSIDIAHTNLKPGKIFIKGNVANYQINRSPSSYLLNPGSERARYSSNTD 180
 Db 145 NFDITCDSIGVQAIYKAHSVQPARMLTQGGELMNSINIRSYAVADNNPEEKAMVYDANVD 204
 QY 181 KEMVLKLVLDNGEDGLISWFAIHPVSMNSNHFVNSDNMGYAALFEQK-KGYLP 239
 Db 205 KMTVIRIEDMSGNPFALISFEFGVCHTSMNTNHLISGDNKGYASYLMEKHANQSSLP 264
 QY 240 QGPVAVAGPSSNLGDVSPNIIIGPHCVNTGESCDNDKSTCPNGGSPMCASGPG--DMFE 297
 Db 265 TGFPIAAGGSGNEGDVSPNTRGPTC-RKGKCDYKTSTC-NGKYVECHALGPGIDGMFE 322
 QY 298 STHTIIGRIYQAKELIYASASOEVTGPVLAHOMVNMIDVSVOL---NATHVTCKPAL 354

Db 323 STIIIGNOFNKLELEFNNAITIOVSGKIQRHTWKPFNTNSVEAPNVSVEGATTCGAM 382
 QY 355 GYSAFAGTIDVSGSLNTIQC-TTEGDFMDTLRDOLGKPSSEIECOKRPILLHSGEL 413
 Db 383 GYSAFAGTIDVSGSLNTIQC-TTEGDFMDTLRDOLGKPSSEIECOKRPILLHSGEL 413
 QY 414 TIHPMOPDIDVQIVTSGSLAIAIPGELTMSGRFRPAIKKEPALYMKDMTIVYIAG 473
 Db 441 VERIPMPVDPMPLOITLQIIVLAVAGEFTTMSGRRLNTV-REIGOSIENPVIYIAG 499
 QY 474 LSNVHTYITTEYEOAROEASTIYGPHTLSAYIOLFRLAKAIAIDVANNSSP-- 531
 Db 500 LAMTISYITTEYEOAROEASTIYGPHTLSAYIOLFRLAKAIAIDVANNSSP-- 531
 QY 532 ---EPPEFKNLIASLIPNIDRAPIGKHFQDVLQPAK--PEYRGEVEYEVFGANPKNS 586
 Db 560 MSCHTPEF-----LPPVIVDAVAPKPFDDIDTDTSTTPSYINQVTVYIIFGANNRN 613
 QY 587 AENQIOTFLTYEKEDSVADQIMNDASMETRFYWHKGLGLSNATYIWHIPDTAYPG 646
 Db 614 F--MTSSSFLTVDOLO-SNGQMTTINDGDMDTKLYMKHDLGFSLITVDMTISPTIOPG 670
 QY 647 IYRIYFGHNRKQELLKPAVILAFEGISSPEFV 679
 Db 671 TYRIHSGVAKKAPFSDN--LTFYQGISNENY 701

RESULT 9

Q9SH86 PRELIMINARY; PRT; 715 AA.
 AC Q9SH86;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DE AT2938010 protein.
 GN AT2938010.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_Taxid=3702;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RX MEDLINE=20083487; Pubmed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eissen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:761-768(1999).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC007661; AAD32770.1; -
 DR InterPro: IPR000418; Ets.
 DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
 SQ SEQUENCE 715 AA; 78505 MW; 075B40D9F9E2AE8B CRC64;

Query Match

Best Local Similarity 37.4%; Score 1249.5; DB 10; Length 715;
 Matches 280; Conservative 113; Mismatches 226; Indels 129; Gaps 18;

QY 4 YTVGSRADCTGVSDINLNGYKNGCNGARGLTFLFSRAFLADPDGSRMAFVSVELC 63
 Db 26 YLVGVSVDITGGAADVNMGMVANSQDIASGIFRLARAFIYAE----- 71

QY 64 MISORLEVLKRLKESKYSGLYRRDNVILSIHTHSGPAGFOYTYIILASEGFSNTEQ 123
 Db 72 -----OENVAISGITHAGPGGLYVTVYIYSLGFRQSPD 109
 QY 124 YIVSGIKSIDIAHTNPKPKIFLNKGNVANNVQINRSPSYLLNPOSERARYSSNDKEM 163
 Db 110 VVANGIEOSTIVQAHESLRPSAFVNNKGDLLDAGVNRSPSSYLNNPAERSKYRYVDKEM 169
 QY 184 LVKLVDLNGEDDLGLSMFAIHPVSNMNSNHFVNSDMGYAAVLFED-----EKNK---- 234
 Db 170 TLVKFVDSQGLPGPSFWMFHTHTSMRTSLISGDCKGAARFMEDFMENQKNSVSR 229
 QY 235 -----GYPGOG-----PRVAGFAS 250
 Db 230 NIPRRVTSVDSERNOSRLDIAATYKSSRGSVDKSLDYKTRVRNKRKKEVSAFCQS 289
 QY 251 NLGDVSNILGPHCVNNGESCDDKSPCPNGCSMCAASPG-QDMESHIIIGRIITYOK 309
 Db 290 NCGDVSPNLTGTCIDTGLPCDFNHSTC-NGONELCYGRGPGYPDEFESTRIIGEKQFM 348
 QY 310 AKELYSASQEVTPVLAHQWVNMVDVSOL-----NATHVTCKPALGYSPAAGTIDG 365
 Db 349 AVELFNKATEKLGKIGYQHAVIDFSNLDVTVKAGGSGSTVTCRAANGFGAAGTIDG 408
 QY 366 VSGLNTIQTGTEGDFMDTLRDOLGKPSSEIECOKRPILLHSGELTIHPMOPDIYD 425
 Db 409 PGAFDFKQGDGQVFWRLVRN-VLRTPGPEQVQCKPKPILLDTGEMKEPEYDM----- 461
 QY 426 VOIIVTGSALIAIPGELTMSGRFRPAIK-----KEPALYMKDMTIVYIAGLSNV 477
 Db 462 -ALIRIGQVILSPGSEFTTMAGRLRLDAIKSEISSDPKEFS-----NNHNVYIAGLTNT 516
 QY 478 YTHYITTEYEOAROEASTIYGPHTLSAYIOLFRLAKAIAIDVANNSSGPEPP-FE 536
 Db 517 YSQYIAFEEYEVQRYGASTIVGRHTLVAYIDGFEKRLATLVNG--LTPRGPQPDLL 574
 QY 537 KNLIASLIPNIDRAPIGKHFQDVLQ--PAKPEYRGEVEYEVFGANPKNSLNDQHOT 594
 Db 575 DKQISLSPVAVVSTPLGVGKVDVPRKSTFRRCQVNAFTWSCGCPN--DLMTGGS 632
 QY 595 FLTYEKEDSVADQIMNDASMETRFYWHK--GIIGLSNATYIWHIPDTAYPGIYRIY 652
 Db 633 FAYETIREG-GKAPRYDDDDSLKFKMSRPKLSSEQATTEMRPESAAGVYIRIRH 691
 QY 653 FGHNRKQELLKPAVILAFEGISSPEFV 680
 Db 692 YGASKSLF-----GSISFSGSSSAFYV 715

RESULT 10

Q9FILA PRELIMINARY; PRT; 705 AA.
 AC Q9FILA;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE Neutral ceramidase.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_Taxid=3702;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=99156233; Pubmed=10048488;
 RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
 RT Sequence features of the regions of 1,081,958 bp covered by seventeen
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:379-391(1998).
 DR EMBL: AB016885; BAB09641.1; -

SQ SEQUENCE 705 AA: 78426 MW: 52C77CD9F823B680 CRC64;
Query Match 34.1%; Score 1223; DB 10; Length 705;
Best Local Similarity 38.1%; Pred. No. 2,9e-88;
Matches 287; Conservative 105; Mismatches 211; Indels 150; Gaps 22;
QY 2 SGYYIGVGRADCTGQVSDINLMGYKNGONARGLITRLEFRATLADPDG-SNRMAFVSVE 61
DB 27 SDYLMGLGSDYITGPAAVDNMMGYANMEQVAGSVFRLRARAFIYAE----- 74
QY 62 LCMISQRLREVLKRLKESKYSLYRRDNVILSAHTHSGPAGFFOYTYLYLASGFSNRT 121
DB 75 -----YKENVAISGTHHAGPGGYLYQYIYLYVSLGFEVHS 110
QY 122 FOYIVSGIMKSIDAHTNLKPGKIFINKGVANVQINSPSSYLLNPOSERRARYSSNDK 181
DB 111 FNAIVLDGIEQSTIQAHENLRPSGILINKGELLDAGVNSPSAYLNNPAHERKRYDYDVK 170
QY 182 EMLVLKLVLDNGEDGLISWFAIHFPVSNMNSNHFVNSDNMGYAALY-----FEOEKNGYL 237
DB 171 EMTLVKFEVD-----DQW---GPVGSFNC-----GNNKGTAAARIMELWFERENGCSKV 214
QY 238 PGQGP-----FV 244
DB 215 DVESFRVSYIISDPYGEHODLMEMASSLSTGCKTVTRMSSVARRVK:RFRHAKPRFV 274
QY 245 AGFASNSLGDVSPNLTGPHCVNTGSGCNDKSTCNGSPSCMASPG-QDMFESTHILIG 303
DB 275 SAFCTNCGDVSPNLTGPHCVNTGSGCNDKSTCNGSPSCMASPG-PDEESTRIIG 333
QY 304 RIYOKAKELYASASQEVTPGLAHOVNMNTDVSVOJN---ATHVTKCPALGYSFA 359
DB 334 EROFKKADLFTKASEEIOGKYDVRYHAYVDSQLEVTINGONGCEVVKTCOPAMGFGFA 393
QY 360 AGTIDVSGVLTGTTGTPDPRWDTLRDQLLKSESEVECOKPRPILHSGELTPIHPW 419
DB 394 AGTIDVSGVLTGTTGTPDPRWDTLRDQLLKSESEVECOKPRPILHSGELTPIHPW 452
QY 420 QPDIYDVQIVTVGSLAIAIPGELTMTSGRRPREIK---KEFALYGMKDMTVYAGISN 476
DB 453 -----ALIRIGOLVILCVBGEFTTMAGRRLRDVAVKTVLKEGS--NKRFSVYIAGLTN 503
QY 477 VYTHYITTEYEQARBAASTIYGPRTLISAVIOLFR---DLAKAIAITDVANMSSGPE 532
DB 504 SYSQIATFEYEQYQRYGASTLYGCPHTLSGYIOEFKLANDLSAQGTD-----PPQ 557
QY 533 PP-EFKNLIASLIPNIADRAPIGKHFGDYLQ--PAKPEYRVG-EVEVEIVGANKPSAE 588
DB 558 PPDLLHKOISLTPVADMTPIGTAFGDTSDVPRLSKFRKGADIVRYQFRSANDRN--D 615
QY 589 NOTHOTFLTVKREYEDSVADQIMYNDASMETRFEVHK--GILGSLNATIIYHIIPATAPG 646
DB 616 LMTSTFALVERWLEGRRTWVPVYDDDFCLRFKMSRPFKLSTOSTATILEMRTETASPG 675
QY 647 IYRIYFEGHNRKQELLKPAVILAEGLSSPEV 679
DB 676 VYRITHFG-SAKTPI---SIIHFSGSSSAFV 704
RESULT 11
Q8X0X9 PRELIMINARY: PRT; 841 AA.
AC Q8X0X9;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Conserved hypothetical protein.
GN 123A4.100.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]

RP SEQUENCE FROM N.A.
RA Schulte U., Algn V., Hobeisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.,
RA Submitted (Jan-2002) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RA Submitted (Jan-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AL670009; CAD21363.1;
KW Hypothetical protein.
SQ SEQUENCE 841 AA: 92901 MW: 0E3E8570314B3A0B CRC64;
Query Match 31.3%; Score 1120.5; DB 3; Length 841;
Best Local Similarity 36.5%; Pred. No. 5.4e-80;
Matches 269; Conservative 119; Mismatches 279; Indels 69; Gaps 23;
QY 4 YIYIGVGRADCTGQVSDINLMGYKNGONARGLITRLEFRATLADPDG-SNRMAFVSVE 62
DB 110 YLGVGKGDITGPVVEINLMGYADPKQGTGLRQRLYSRAFTVSLERPODFVYLVDT 169
QY 63 CMISQRLR---LEVLRLESKYSLYRRDNVILSAHTHSGPAGFFOYTYLYLASGFSN 119
DB 170 QSGDTAVRFGILKALKELGPEY-APYGHNNALTGTHSHAGGGLNLTLLQITSGFDR 228
QY 120 RTFOYIVSGIMKSIDAHTNLKPGKIFINKGVANVQINSPSSYLLNPOSERRARYSSNDK 175
DB 229 QCYQATVGVAGVILSKAHESLOPYLSAGTTKVFGANINRSLFSLANPEARMKTYFSE 288
QY 176 --SSNTDKEMVLKLVLDI-NGEDGLISWFAIHFPVSNMNSNHFVNSDNMGYAALYFEOEK 232
DB 289 EEDGSVEKDLITMLKFORASDCKNIGVITWPFHGTSMIGNTVYGDKNGVAAWLFESV 348
QY 223 NKGYLPGGPFPAGASSNLGDVSPNLTGPHCV-NTGSCNDKSTCNGSPSCMASPG 291
DB 349 -RGDSSAAEDFVAGSQANMGVSPNLTGAWCEDSGQCSFKNSTCNGSQACHGRP 407
QY 292 G---QDMFESTHILIGRIYOKAKELY---ASASQEVTP-GLAHOVNMNTDVSVOJN 343
DB 468 AFKVHDDGASSGFEIGRQFEPKAKOLYHLLNNSPVGPMYKATHTHDMNSNFTPIRPD 467
QY 344 THVYTKCPALGYSFAAGTTDGVSGSLNITQ---GTEGDPFMDTLRDQLKPRSEIV 398
DB 468 GKPARTCPALGYSFAAGTTDGPAGFDFTOHNGNENTYM--PIKVVVR-YFIKAPGDQK 524
QY 399 ECKQKPRILLNSGELTTPHPOPDIVQIVTVGSLAIAIPGELTMTSGRRPREAIKE 458
DB 525 ACQHPKPLILDVGETLRPDMSPNIVQAFRFGQFALVVSFGAFTYAGRRMKEAVGQS 584
QY 459 FALYGMKDM-----TVYIAGLSNVYTHYITTEYEQARBAASTIYGPRT 504
DB 585 FKRLIDQDMQETGYGTGPPSSNPVYLGGRANTYHTTTEBEYIOQRYEGASTLYGPD 644
QY 505 LSAVITQFLRDLAKAIAITDVANMSSGPE--PPEFKNLIASLIPNIA-DRAPIGKHFGDYL 561
DB 645 LNAVYNTVLSFLPYLSESTSPRAHDEKSPDPDNRSLSPFAYVRDAPPLFKFGDYL 704
QY 562 QPAKE--YRGEVVEVIFVGANPKNSAENQTHQELVYK-YEDSV-----ADMOIMND 614
DB 705 VDVDSARGYHRGDALRAVFGANPRNNL--RLEGTYAAVEKLFQDVQVPOKSEMTVYSD 762
QY 615 ASMETRFEVHK--GILGSLNATIIYHIIPDTA-----YGIYIRIRYFGRNRKQELLKP 664
DB 763 EDMSLITMYKRNKNSVNGTSVEIIMETGEETDEMDKDELGCYVLYLKYGDSKS---LFG 819
QY 665 AVIILAEGLSSPEV 680
DB 820 GKVOEFEGVSGAFTLY 835
RESULT 12
Q9I596 PRELIMINARY: PRT; 670 AA.
AC Q9I596;
Q9I596;

DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Hypothetical protein PA0845 (Ceramide-hydrolysing enzyme precursor).
 GN PA0845.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 NC NCB1_Taxid=287;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15592 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith J.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 RN (2)
 RP SEQUENCE FROM N.A.
 RA Nieuwenhuizen W.F.;
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RA Okino N.;
 RT "Molecular cloning, sequencing, and expression of the gene encoding
 RT alkaline ceramidase from Pseudomonas aeruginosa. Cloning of a
 RT ceramidase homologue from Mycobacterium tuberculosis.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE004519; AAG04234.1;
 DR EMBL: AJ15932; CAG67511.1;
 FT HYPOHETICAL protein: Signal: Complete proteome.
 FT SIGNAL 1 26
 FT CHAIN 27 670 CERAMIDASE.
 SQ SEQUENCE 670 AA: 73372 MW: 712073EAC9CED287 CRC64;
 Query Match 30.4%; Score 1095.5; DB 16; Length 670;
 Best local Similarity 35.9%; Pred. No. 3.6e-78;
 Matches 250; Conservative 119; Mismatches 253; Indels 75; Gaps 14;
 QY 4 YIIGVGRADCTGQVSDINLMGYKNGQONARGLTLRFSAFIADDPGSRNAFVSELC 63
 DB 29 YRGLGKADITGGAEEVGMVSSLEOKTAGIHRQMARAFVIEEASGRLLVYNTDGL 88
 QY 64 MISQRLLEYKRLKESKYSGLYRRDNVILSAITHSGPAGFFQYTYILASEGFSNRTQ 123
 DB 89 MTFQAVHLKYLAKKPYGVYDENNVMLAATHHSGPGGFSHYAMNLSVLGFOEKTEN 148
 QY 124 YIVSGIMKSIDIAHTNLKPKGIFINKGNVANOINSPSSYLINPOSERARVSSNTDKEM 183
 DB 149 AIVDGIVRSIERQAALQPGRLFYGGSELNARNRNSLSHLKNP--DIAGYEDGIDPOM 206
 QY 184 LVKLVDLNGEDGLISWFAIHPVSMNNSHHFVNSDNMGAAVLFQOEKNKGYLPGQGP 243
 DB 207 SVLSFVDANGELAGALISWFPVHSTMTNANHLISPNKGYSYHWEHDVSR-----KSGF 261
 QY 244 VAGFASNLGDVSPNLTGPHCVNTGSCDNKSTCPNGSPKMGASGPGQDMESHTIIG 303
 DB 262 VAFAFQTNAGNLSPNL-----NLKPGSGPDNEPDNTRIG 297
 QY 304 RIITYQAKELIYASQSEVTGPVLAHQWVMTDVSQVLANAT--HTVKTCKPALGYSFAAG 361
 DB 298 LRQFAKAYEIAQAOEVEIGELDSRRFVDFTLPIRPETDQPOLCTAAGTSLAAG 357
 QY 362 -TIDGVSGLINTGTTGEGDFMDTLRDQLGKRSSEIIVCQKPKPILHSGELTIRHPWQ 420
 DB 358 SFEDGGPIGLEEG---NNPFLSALGGLTGVPPQLVCOQAKETILTADTGNKK--PYPMW 413

QY 421 PDIVDQIYTVGSLAIAALPGLTMTSGRRFRPAIKKEPALGKMDTVYIAGLSNVYH 480
 DB 414 PVLPIOMFRIQIGLELIGAPAEFTVAGVIRRAVOAASPAEIRH--VVENGYANAYAS 471
 QY 481 YITTEGYOQAREASTIYGPHTLSAIOLEFDLAKA-----INTDVA-----N 526
 DB 472 YITTEGYOQAREASTIYGPHTLSAIOLEFDLAKA-----INTDVA-----N 526
 QY 527 MSGPPEPPEKFNLIASIPINADRAPIGKFGDVLQPAKREYVGEVEYIEVGANKPS 586
 DB 532 FQTV-----VADDPYIGKSGFDVLDQPPRESYIGKVTYAFYTGHPKN- 575
 QY 587 AENQHOTFLTYKTY-EDSYADQWIMYNDASWETRTYMHKGLIGLSNATIIYHIDPTANP 645
 DB 576 -DIRTEKTFLEVYVNIKKQKQPEYATVATDMDQTOYRWERVGVISAKATISIPGTEP 634
 QY 646 GYIRYFPGHNRKQELLKPAVILAIFGSSPEFVYTT 682
 DB 635 GHYIRHYGNAKNFYTK---ISEIGSTRSEFVLT 668
 RESULT 13
 Q9RHO0 PRELIMINARY; PRT; 670 AA.
 AC Q9RHO0;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE Alkaline ceramidase.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 NC NCB1_Taxid=287;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20062886; PubMed=10593963;
 RA Okino N., Ichinose S., Omori A., Imaiya S., Nakamura T., Ito M.;
 RT "Molecular cloning, sequencing, and expression of the gene encoding
 RT alkaline ceramidase from Pseudomonas aeruginosa: Cloning of a
 RT ceramidase homologue from Mycobacterium tuberculosis.";
 RL J. Biol. Chem. 274:36616-36622(1999).
 DR EMBL: AB028646; BAA8409.1;
 SQ SEQUENCE 670 AA: 73275 MW: 3E9FBCB3521AC0 CRC64;
 Query Match 30.4%; Score 1090.5; DB 2; Length 670;
 Best local Similarity 35.9%; Pred. No. 8.9e-78;
 Matches 250; Conservative 118; Mismatches 254; Indels 75; Gaps 14;
 QY 4 YIIGVGRADCTGQVSDINLMGYKNGQONARGLTLRFSAFIADDPGSRNAFVSELC 63
 DB 29 YRGLGKADITGGAEEVGMVSSLEOKTAGIHRQMARAFVIEEASGRLLVYNTDGL 88
 QY 64 MISQRLLEYKRLKESKYSGLYRRDNVILSAITHSGPAGFFQYTYILASEGFSNRTQ 123
 DB 89 MTFQAVHLKYLAKKPYGVYDENNVMLAATHHSGPGGFSHYAMNLSVLGFOEKTEN 148
 QY 124 YIVSGIMKSIDIAHTNLKPKGIFINKGNVANOINSPSSYLINPOSERARVSSNTDKEM 183
 DB 149 AIVDGIVRSIERQAALQPGRLFYGGSELNARNRNSLSHLKNP--DIAGYEDGIDPOM 206
 QY 184 LVKLVDLNGEDGLISWFAIHPVSMNNSHHFVNSDNMGAAVLFQOEKNKGYLPGQGP 243
 DB 207 SVLSFVDANGELAGALISWFPVHSTMTNANHLISPNKGYSYHWEHDVSR-----KSGF 261
 QY 244 VAGFASNLGDVSPNLTGPHCVNTGSCDNKSTCPNGSPKMGASGPGQDMESHTIIG 303
 DB 262 VAFAFQTNAGNLSPNL-----NLKPGSGPDNEPDNTRIG 297
 QY 304 RIITYQAKELIYASQSEVTGPVLAHQWVMTDVSQVLANAT--HTVKTCKPALGYSFAAG 361
 DB 298 LRQFAKAYEIAQAOEVEIGELDSRRFVDFTLPIRPETDQPOLCTAAGTSLAAG 357
 QY 362 -TIDGVSGLINTGTTGEGDFMDTLRDQLGKRSSEIIVCQKPKPILHSGELTIRHPWQ 420


```

Db 358 STEGPGPLGLEEG---NNPFLSALGILTCVPPPELVOCOAEKILALATGNKK--PYWT 413
OY 421 PDIVDOVLVYVYGLSLAIAIPGELITMGSRRREAIKKFALYGMKDMVYVAGLSNYTH 480
Db 414 PTVLPIQWFRIGQLLELGAPEFVMAVIRRAVAQASSEAAGIRH--VVENGVANAYAS 471
OY 481 YTTVEEYOAOYEASTTYGPHLTSAVITQLFRDLAKA-----IATDYVA-----N 526
Db 472 YVTRREEAQAEYEGSTLYCPWTOAAQOLFVDMAVALLRERLPVETISNAPDLSCQMN 531
OY 527 MSSGPEPPEFKNLLASLIPNADRAPIGKHGFDVLOQAPKPEYRGEVVEVTFVGANPKNS 586
Db 532 FQTVGV-----VADDPYIKSGFDVLOQPRESTRIDKVTVAFTGHPKRN- 575
OY 587 AENQTHQTLVEXY-EDSVADMQIMYNDASWETRFYHKKGLIGSNATYIWHIPDTAYP 645
Db 576 -DLTEKTELEVNIQKQKOTPVATNDMDQYMERERGISASKATISMISIPGTEP 634
OY 646 GYIRIRYEGHNRKQELLKPAVILAFEGISSPEEVTT 682
Db 635 GHYIRHGNKNAKNFWTK---ISELGGSTRSEFVLGT 668

```

RESULT 14

006769 PRELIMINARY: PRT: 637 AA.

```

ID 006769
AC 006769:
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
GN RV0669C OR MTC1376.05.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Galloway S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL: Z95972; CAB09388.1;
DR TubercuList; RV0669C;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 637 AA; 69489 MW; 5E2D915AB1E4FDB7 CRC64;

```

Query Match 24.7%; Score 884.5; DB 16; Length 637;
 Best Local Similarity 32.3%; Pred. No. 1.9e-61;
 Matches 221; Conservative 113; Mismatches 291; Indels 59; Gaps 15;

```

OY 6 IGVRADCTGVSDINILMGYCKNGNAGLTLRLFSRAFIAD--PDGSNNMAEFSVELC 63
Db 4 VGRGIADITGEADACMGYCKSPQRTAGIHQRLSRAFRVDRDSQDDADALLILVALKP 63
OY 64 MISORLLEVLKREKSYGSLYRDNDVLSAIFHSGPAGFPQYTLYLASEGFSNRTFQ 123
Db 64 LPMQNVNEEVLRRLADYDGYSEONLTITATHTHAGGCGYLYNMLTSSGFRPATFA 123
OY 124 YVSGIMKSIDIAHNLKPKGIFINKGNVAVNOINRSPSSVLLNQGSRARVSSDTKEM 183
Db 124 AIVQIVESVHAHADVAPEVSLSHGELYGASINRSPSAFDRNPIADKAPFPKRVDPHT 183

```

```

OY 184 LVLLKVLNDGEDLGLISWFAIHPVSMNNSNHFVNSDNNGAYAYLEOE-KNKGYLPGGCP 242
Db 184 TLVR-IDRGEATVGYIHFFATGHTSMTRNHLISDNKGPAAYHMERIVGADYLAQGPD 242
OY 243 FVAGASSNLGDSVSNILGPHCVNTGESCDDNDKSTCPNGGFSMCASPGQDMFESTHII 302
Db 243 FIAPFAQTNPGDMSNVNDGP-----LSPEAPPPREPTNTRT 279
OY 303 GRIIYOKKELAYASQEVTPGLAAHQVMNMTDVSVOLNAT---HYKTKCPALGYSFA 359
Db 280 GLCOPEFAPFQLSGATPIGAG-IDARFTYVDLSVLYGERTPDGCEERTGCRPMGAGAM 338
OY 360 AGTIDVSGSLNTQGTGDPFMDPLRDOL--LGRSESEIVECQKPKPILHSGELTIPH 417
Db 339 AGTDGCPGFHGRQCR---NFPWDLSSAMRIANFT---AAQAPKQIVNPARLPNRIH 392
OY 418 PMQPDIVQIVYVYGLSLAIAIPGELITMGSRRREAIKKFALYGMKDMVYVAGLSNV 477
Db 393 PFVQEIYVQVQVIRIGRLYLIGIPGPTIVAGIRLRMYA---SIVGADLADVLCGYTNA 449
OY 478 YTHYTTVEEYOAOYEASTTYGPHLTSAVITQLFRDLAKAIAVDTVANMSSGPEPPEK 537
Db 450 YIHVYTPPEETLEQRTBEGSTLFGWELCALMQYVAELAEAMRDCRPVTLGRPPRTREL 509
OY 538 NLIASLIPNIDRAPI-GKHFGDVLOQAPKPEYRGEVVEVTFVGANPKNSAENQTHQTEL 596
Db 510 SWVRG-----ANADAGSFGAVIAEESATYRPGQAEAVFVSALPNN--DLRRCGYTL 559
OY 597 TVEKEDSVADMQIMYNDASWETRFYHKKGLIGSNATYIWHIPDTAYPGIYRIRYEGHN 656
Db 560 EVVRREG--ASWVRIADGDGMATSPRMORGRAGSHVISIRMDVPGDTYRIVHGSTA 617
OY 657 RKQELLKPAVILAFEGISSPEEVY 680
Db 618 RDRN---GMLTASATTRFTTV 637

```

RESULT 15

093216 PRELIMINARY: PRT: 314 AA.

```

ID 093216
AC 093216:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE AT558980/K19m22.180.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC Cheuk R., Chen H., Kim C.J., Koeseema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamuya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsids cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY057506; AAL09747.1;
SQ SEQUENCE 314 AA; 35401 MW; 6FE09DA6B8FF7434 CRC64;

```

Query Match 17.5%; Score 626; DB 10; Length 314;
 Best Local Similarity 45.6%; Pred. No. 2.1e-41;
 Matches 144; Conservative 48; Mismatches 96; Indels 28; Gaps 12;

```

OY 377 EGDPPWDLRLQOLLGSPSEIYECQKPKPIILHSGELTIPHPMODIVDQIVYVYGLAI 436
Db 13 KNPFWRLVRN-LKNPTEBOVRCORPKPIILDTEGMRQPYDMASVILPVOILRIGQLVI 71

```


DB 335 KRC---FSAFCOTNGDVSPNVLGAFCLDTGLPCDFNISTC--GKNEKMGYRGPGPDE 390
QY 296 FESHTIIIGRIYOKAKELASASQEVTPGLAHOVNMMDVSVOL-----NATHVTCK 351
DB 391 FESRTRIGERQFMALELEFNKASEQLOGKVDYRHVYVDSQLVTLPKKCKSEYVKTCP 450
QY 352 PALGYSFAGTIDVSGGLNTGCTTGSDPFMTLRLDQLGKPSSEIVECOKPRPILHSG 411
DB 451 AAMGFAPAGTTDGPAPFETQDDKGNPFWRJVRN-VLKTDPKOIDCHYKPIILDTG 509
QY 412 ELTIPHPMDIYDVQIVMGSLAIAIPGLTMSGRREARAIKKEFALYKMDMT--- 468
DB 510 EMKRPIDMAPSILSLDYLRIQLFLISVEEFTTMAGRRLRYAVKTKLNSGKNKSDGET 569
QY 469 -VVIAGLSNYTHYTTVEEYQARYEAATSTYGPHTLSAYIOLFRDLAKAATDVTYAM 527
DB 570 HVIYAGLANGSQTVEEYQVREGASTLYGPHTLSCYIOEPKLSLVLDD--MPV 627
QY 528 SSGPEPP-FFKNLIASLIRIADRIAPIGHFQVLO--PAKPEYRGE-VVEYIEGANG 583
DB 628 QPQPQPPDLLDKQLSLTPVMDTPSGDSFGDIVSDVPKLSLKGNGQVTVVFRSACP 687
QY 584 KNSAEQHTQTFLETVEKEDSVADMOIMYNDASMETREYV--HGILGSLNATIYHHPD 641
DB 688 RN--DLTGETFLVLERLQKDKTTPVYDDDLCLRFKRSRKKLSNSQATVENRIPE 745
QY 642 TAVPGYRIRYEGHNK-----QELLKPAVILAFEGISSPFEV 679
DB 746 SASPGYRITHFGCAKKLFGQVRQNGVKELMRSLMESNIPFHV 789

RESULT 2

H84799
hypothetical protein AC2938010 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: H84799
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shua, T.P.; Beutlo, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Gues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:1061197
A:Accession: H84799
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-715 <STO>
A:Cross-references: GB:AE002093; NID:g4895183; PIDN:AMD32770.1; GSPDB:GN00139
A:Gene: AC2938010
A:Map position: 2
C:Superfamily: Mycobacterium tuberculosis hypothetical protein RV0669c

Query Match 34.9%; Score 1249.5; DB 2: Length 715;
Best Local Similarity 37.4%; Pred. No. 9.9e-85;
Matches 280; Conservative 113; Mismatches 226; Indels 129; Gaps 18;
QY 4 YIYGVGRADCTGVSDINIMGYGKNGONARGLTLRLFSRAFLIADPDGSRMAFVSELC 63
DB 26 YIIGVGSYDTGTGAADVNMGMGYANSQDIASGIFRLRARAFIYAE----- 71
QY 64 MISQRLREYLKRLSKYSGSLYRRDNYLSAITHSGPAGFOYTYLILASGFSNRTQ 123
DB 72 -----OEKNVAISGITHAGPGGYLQYTVYIVYISLGFVRSQSP 109
QY 124 YIYSGIMKSIDIAHTNLKPKGKIFINKGNVANOINSPPSYLLNPOSEARARYSNDKEM 183
DB 110 VVANGESQSIYVQHESLRPGSAFVNGDLDGAVNPSPPSYLLNPAERSKYVYVDKEM 169
QY 184 LVLLKLDLNGEDLGLISWFAIHPVSNMNSNHFEVNSDNMGYAAVLFEQ-----EKNK----- 234
DB 170 TLVKKFVDSQLGPTGSENMWATHTGTSKRSINSLISGKNKAAAFMEDWENGNKNSVSR 229

QY 235 -----GYLPQG-----PVGAFSS 250
DB 230 NIPRNVSTIVDSERNQSRLLDIAATYKSSRGHSVDKSLDVKTRVNGSKRKFEVAFCS 289
QY 251 NIGDVSPTNLGPRHCVMTGESCDNDKSTCPNGSGSMCASGPG-QDMFESHIIIGRIYOK 309
DB 290 NCGDVSPTNLGTCIDTGLPCDFNISTC--NGONELCYGRGPGYPDEFESTRILIGKOFK 348
QY 310 AKELYSASQEVTPGLAHOVNMMDVSVOL-----NATHVTCKPALGYSFAAGTIDG 365
DB 349 AVELFKNATKLOGKIGYQIAYLDFSNLDVTYKAGGSETVATCAAMGFGAACTTDS 408
QY 366 VSLNTGCTGSDPFMTLRLDQLGKPSSEIVECOKPRPILHSELTPHPMDIYD 425
DB 409 PGAFDEKQGDQGNVFWRLVRN-VLRTPGEQVQCCPKPILDTGEMKEPYDW----- 461
QY 426 VOIVTIGSLAIAIPGLTMSGRREARAIKKEFALYKMDMT--- 477
DB 462 -ALIRIGQLVILSVPGFTTMAGRRLDALKSLISSDKRES-----NNHVIYAGLINT 516
QY 478 YTHYITTYEYQARYEAATSTYGPHTLSAYIOLFRDLAKAATDVTYAMSSGPEPP-FF 536
DB 517 YSQYIATFEYEYQVREGASTLYGPHTLAYIQEPKALATLVNG--LTLPRGQPPDLL 574
QY 537 KNLIASLIRIADRIAPIGHFQVLO--PAKPEYRGEYEVYFVGANKNSAENDTHT 594
DB 575 DKQISLSPVYVDSPLGVYFGVADVPKSTFRGQOVNATFWSGCPRN--DLMTES 632
QY 595 FLTVKEEDSVADMOIMYNDASMETREYVHK--GIILGSLNATIYHHPDPRYGIYRIR 652
DB 633 FAVETLRGCG-GMAVYVYDDDFSLFKNSRPKLSSESQATTEMKVPESAAGVYRIRH 691
QY 653 FGHNRKQELKPAVILAFEGISSPFEV 680
DB 692 YGASKSLF-----GSISFSGSSSAFYV 715

RESULT 3

C83540
conserved hypothetical protein PA0845 [imported] - Pseudomonas aeruginosa (strain PAO
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83540
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82850; MUID:20437337; PMID:10984043
A:Accession: C83540
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-670 <STO>
A:Cross-references: GB:AE004519; GB:AE004091; NID:g9946736; PIDN:AMG04234.1; GSPDB:GN
A:Experimental source: strain PA01
A:Gene: PA0845
C:Superfamily: Mycobacterium tuberculosis hypothetical protein RV0669c

Query Match 30.6%; Score 1095.5; DB 2: Length 670;
Best Local Similarity 35.9%; Pred. No. 2.6e-73;
Matches 250; Conservative 119; Mismatches 253; Indels 75; Gaps 14;
QY 4 YIYGVGRADCTGVSDINIMGYGKNGONARGLTLRLFSRAFLIADPDGSRMAFVSELC 63
DB 29 YRGGLKADITGAAEVMGYSLSQKAGIMRQMARAFIYEAAAGRLVYVTDLG 88
QY 64 MISQRLREYLKRLSKYSGSLYRRDNYLSAITHSGPAGFOYTYLILASGFSNRTQ 123
DB 89 MIFOAVHLKVLAKKAYGCVYDENVMMLAATHSGPGGFSHYAMYNISVLGFOKFTEN 148
QY 124 YIYSGIMKSIDIAHTNLKPKGKIFINKGNVANOINSPPSYLLNPOSEARARYSNDKEM 183

Db 149 AIVGIVRSIRRAQARLPGRFLFYSGCELFRANRRSLSHLKNP--LIVGDEIDIDPM 206
184 LVKLIVDNGEDDLGISWFALHPVSMNNSNFVSDNNGYAAYLEQKKNGYLPQGP 243
Db 207 SVLSFVDNGLAGLISWFPVSTMTNANHLISDNNGYASYHHEHIVSR---KSGF 261
Qy 244 VAGFASSNLGADVSPNLLGPHCVNTGESCDNDKSTCPNGSPMCMAAGGOMFESTHILG 303
Db 262 VAAFAQTNAGLSRNL-----NLKRGSGFDFNDFDTRIG 297
Qy 304 RIYOKAKELYSASQEVTPVLAHQMNTDVSVOLNAT--HTVTKCKPALGYSPAAG 361
Db 298 LROPAKAEIAGQAOEVELGELDSRFVDFTRLPPIREFTDGQRCQTAIGSLAAG 357
Qy 362 -TIDVSGSLNTGTTGTEDEPFWDLRDOLLKPSSEIVECOKPKPILHSGELTIPHWO 420
Db 358 STEGPGPLGLEEG---NNPFLSALGGLTGVPPELVOCAEKTILADTGNKK--PYEWT 413
Qy 421 PDIVDVOIVTVGSLAIAIPGELTMSGRPREAIKKEFALYGMKDMTVIAGLSNVTYH 480
Db 414 PTVLPIDQFRIGOLELGAPEFTYVAGVIRRAVOAASEAAGLTH--VFNNGIANYAS 471
Qy 481 YITTYEYOQRYEASTIYGPHTLSAYTOLFRLDAAK-----IATDTVA-----N 526
Db 472 YVTTFREYAQOYEGSGSTLYGPMWQAQOLFVDMAYALRERLPEVTSALIPDLSCGMN 531
Qy 527 MSSGPPEPFKNLALSLIPNADRAPGKHGVDYLOAPKPEYRGEVEVYFVGANPNNS 586
Db 532 FQTVY-----VADDPYIGKSGVDYLOOPRESYRIGDYATVAFVTCGPKN- 575
Qy 587 AENOTHOTFLTVERY-EDSVADMOIMYNDASMETREFYHKGILGNSNTIYWHIPDTAYP 645
Db 576 -DLTEKTFLELVNIGDKQKTPPEVATDNDMDYRKREYRGISASATISMSLPPTEP 634
Qy 646 GIVRIIRYFGHNRKQELLKPAVILAFEGISSPEFVYTT 682
Db 635 GHYIRHYGNNAKNFWTK---ISEIGSTRSFEVLYGT 668

RESULT 4

H70535
hypothetical protein RV0669c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70535
R: Cole, S.T.; Brosch, P.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID:96295987; PMID:9634230
A: Accession: H70535
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-637 <COL>
A: Cross-references: GB:295972; GB:AL123456; NID:93261790; PIDN: CAB09388.1; PID: g2143290
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: RV0669c
C: Superfamily: Mycobacterium tuberculosis hypothetical protein RV0669c

Query Match 24.7%; Score 884.5; DB 2: Length 637;
Best Local Similarity 32.3%; Pred. No. 1.2e-57;
Matches 221; Conservative 113; Mismatches 291; Indels 59; Gaps 15;
Qy 6 IGVRADCTGVSDINILMGKNGCONARGLTRFSRAFLAD--PDGSNMAFVSELC 63
Db 4 VGRGIADTGAACGMLGYSKQRTAGIHQRLSRKAFVRDSDQDDARLLIYALP 63
Qy 64 MISQRLVLRLESKYGLYRDNVLSAIIHSHGPAFFOYTYLTIASEGFSNRFTQ 123

Db 64 LPMQVNEEVLRLADLYGDTYSQONTLITATHTHAGGGYCYLLNLTSGFRPATFA 123
Qy 124 YIVSGIMKSIDAHNLNPKGIKFIKKGVANVOIRSSSYLLNQSRFATSSYTKDEM 183
Db 124 AIVDGVSEVBEHADVAIPAELVSLSHGELGASINRSPSADRNPPAKAFPPKRVDPHT 183
Qy 184 LVKLIVDNGEDDLGISWFALHPVSMNNSNFVSDNNGYAAYLEQKKNGYLPQGP 242
Db 184 TLVR-IDRGEATVGYIHFFPAHGHGSMTRNHLISDNNGFAPAIHHERIVGADYLAQPD 242
Qy 243 FVAGFASSNLGADVSPNLLGPHCVNTGESCDNDKSTCPNGSPMCMAAGGOMFESTHIL 302
Db 243 FLAFAQYNPDGMSPNVDP-----LSPEAPPDREFDTRT 279
Qy 303 GRITVOKAKELYSASQEVTPVLAHQMNTDVSVOLNAT--HTVTKCKPALGYSPA 359
Db 280 GLCOFEAPFQOLSGATPIGAG-IDARFTYVDLSVLYGVEYTPDGEERTGPRMGAGAM 338
Qy 360 AGTIDVAGSLNTGTTGTEDEPFWDLRDOLLKPSSEIVECOKPKPILHSGELTIPH 417
Db 339 AGTDGPGPFHGRGCR---NPPMDRLSMTAMTRAPF---AAQAPKIVMPARLPNRIH 392
Qy 418 PMQPDIVDVOIVTVGSLAIAIPGELTMSGRPREAIKKEFALYGMKDMTVIAGLSNV 477
Db 393 PFVGEIVPVQIVRIGRLYLIGIPGPTIVAGRLRRVVA---SIVGADLADVLGVYNA 449
Qy 478 YTHITTYEYOQRYEASTIYGPHTLSAYTOLFRLDAAKIAIDTVANMSSGPPEPFK 537
Db 450 YIHVYTFPEEYLEORYEGSSTLYGPMWQAQOLFVDMAYALRERLPEVTSALIPDLSCGMN 531
Qy 538 NLIASLIPNADRAPGKHGVDYLOAPKPEYRGEVEVYFVGANPNNSAENOTHOTFL 596
Db 510 SWVRG-----APADGSGFAGVIAEASATYRQGAIVEAFVSLPNN--DLRGGTYL 559
Qy 597 TVEKYEDSVADMOIMYNDASMETREFYHKGILGNSNTIYWHIPDTAYPGLIRYFGH 656
Db 560 EVVRREG--ASWVRVIAADGDWATSERMORGRAGSHVIRMDVPDGTTPGQYRIYHNGTA 617
Qy 657 RKQELLKPAVILAFEGISSPEFVY 680
Db 618 RDRN---GMLTAFSATREFTV 637

RESULT 5

A11489
Probable peptidoglycan bound protein (LPXTG motif) l1n0457 [imported] - Listeria inno
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: A11489
R: Glasner, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A: Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
Ok, C.; Schuete, T.; Simoes, N.; Tjelle, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A: Title: Comparative genomics of Listeria species.
A: Reference number: AB1077; MUID:21537279; PMID:11679669
A: Accession: A11489
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-2013 <GLA>
A: Cross-references: GB:AL592022; PIDN: CAC95689.1; PID: g16412898; GSPDB: GN00178
A: Experimental source: strain Clp11262
C: Genetics:
A: Gene: l1n0457

Query Match 3.7%; Score 131.5; DB 2: Length 2013;
Best Local Similarity 19.8%; Pred. No. 0.92;
Matches 147; Conservative 95; Mismatches 266; Indels 235; Gaps 37;
Qy 44 FLADPDGSNMAF-----VSEVLCMISQRLVLRLEKYSGLYR--RDNVYLSAI 95
Db 102 YLASQTSKYELELKEGDKVOAKLDLIVKNVDEEVAKTVKSNQLLRSSISDRLFLQA- 160

QY 96 HHSGPAGFFQYTLVLLASEGFSNRTFOYIVSGIMKSIDIAHNLKPGKIFINKGVANY 155
| : : : : :
Db 161 --DSSKATLANYTEQI-----TFNSINFLDGS-----TTLKNGKLVIDFNN-SNL 203
| : : : : :
QY 156 QINRSSSYLLNPOSERARYSSNTDKEMLVKLYD--LNGE--DLGLI--SMFAIRPVS 209
| : : : : :
Db 204 ELVNYPKDYAANNINISTYSALTGK--LTINLVNIDISSGAPPDIPVAGARGVP 261
| : : : : :
QY 210 NNSNHFVNSDNMGYAY-----LFEQKNKGYLPQGGPFVAG-----246
| : : : : :
Db 262 MNKATLSEGENSGATYTAKEKTYVLENSNDY-----SPITAGDNSMARSMEELAR 317
| : : : : :
QY 247 -----FASSNLGD--VSPNIILOPHCVNTGESCDNDK 275
| : : : : :
Db 318 SLKPGGYTIQWPEIQKSSSEKSKMLKLEFLKENGDDISVNTADPYVIRFGEPIWSQL 377
| : : : : :
QY 276 STCPNGSPSCMASGCGDMFE-----STHIGRIITQKAKELYASA--SEVYGPVLAH 329
| : : : : :
Db 378 STV--NGKANVLDNDEKQVEYGPINANTYQRIQVSMAPKIPADAVKGTETGYVNYD 435
| : : : : :
QY 330 QWVNMIVSVQNLNATHTVTKCPALGYSPAGTIDG--VSGLNTGCTGEGDPFMDLRQ 388
| : : : : :
Db 436 EDLVNYSIKITKEVTDSATSI-----AVDSKVSSTISGVDYLEMGMPRISSA 464
| : : : : :
QY 389 LIGKPESEIVECKPRPILL-----HSGELTIPHP-----WOPDIY 424
| : : : : :
Db 485 AFGVNDLEIV-APIPKIKVLSYIPNNNSMASMKLEYQNGKWSMAPQTSGMPFSKI 543
| : : : : :
QY 425 DVQIVTVGSLAIAIRGELT-----TMSGRFRAIKKEKALGMKD 466
| : : : : :
Db 544 DQSVNRIEKIKLRSGLIINDKMDPYTHGTIRMONTGVKAGESF--TLQPESTITDSD 601
| : : : : :
QY 467 MFWVIAGLSNVYHTTYTEEYOQRYEAST-----YCPHLSAYIOL 511
| : : : : :
Db 602 KT-----SKAIDTASVYEK--KVQVEKTSITPAKINGDVELSTAGIYCKGESEI--I 651
| : : : : :
QY 512 FRDLAKAINTDVANNSGDEPPEFKNLASLIPNADRAPICKHFGDVLQPAKPEYRVG 571
| : : : : :
Db 652 FFGNDKIAQSVRLSGYSKLENPY-----IFVVPKCIDVETM--KNF---YR-- 698
| : : : : :
QY 572 EYVEYFVGAN-----PKNSAENQTHQFTLYEKEDSVADQWIMVNDASWETRYMHK 625
| : : : : :
Db 699 SIINTYVAPANGNTNLTYPKSSAD-----VKKRETLSDGSTLY-----735
| : : : : :
QY 626 GILGLSNATYWHIPDTAY-PGI 647
| : : : : :
Db 736 -----YWEAPDTGLAPGM 748
| : : : : :
RESULT 6
AF0263
protease IV (EC 3.4.21.-) [Imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: AF0263
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Parraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; M01D:214/0413; PMID:11586360
A:Accession: AF0263
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-616 <KUR>
A:Cross-references: GB:AL590842; PIDN:CA090970.1; PID:q15980166; GSPDB:GN00175
C:Genetics:
A:Gene: sppA
C:Superfamily: protease IV
C:Keywords: hydrolase; serine protease

Query Match 3.7%; Score 131; DB 2; Length 616;

Best Local Similarity 20.7%; Pred. No. 0.15;
Matches 136; Conservative 87; Mismatches 223; Indels 210; Gaps 33;
QY 102 AGFOYTLVLLASEGFSNRTFOYIVSGIMKSIDIAHNL-----KP-----GKIFINKG 150
| : : : : :
Db 9 AGFFKMTWRL-----NFTRELINFLALLILIGVGIYFOROSKPEVPEYKALLVNS 62
| : : : : :
QY 151 NVANVO--IN--RSPSSYLLNPOSERARYS-----MTDEMLVKLVLDN 192
| : : : : :
Db 63 GVIVDPALINNKLRQNGRELLGASSNRLQNSLFDIVETIRLAKDDDNINGVLISDLT 122
| : : : : :
QY 193 GEGLGLISWF--AIHVSNNNSNHFVNSDNMGAAALFEDEKKGKLPQGGP--VAGFAS 249
| : : : : :
Db 123 GADQSSLYQYGMKLRFRDTRKRIYAVGSYNOTOYLLASFARKIYLSPGAVDLHGFA 182
| : : : : :
QY 250 SNLDGVSPIILGPHCVNTGESCDNDKSTCPNGSPSCMASGPGDMFESETHIIGRIYQK 309
| : : : : :
Db 183 NNL-----YYSLEEN-----LKVTNIRVGTYS 208
| : : : : :
QY 310 AKEL-----YASASQEVTPVLAHQVN-----MTDVSVOLNATHTVTKCPALGYSP 358
| : : : : :
Db 209 AVEPMIRNDMSAARE-----ADSRVGGLMQNYLTIVSANRLT-----PEQLFPG 255
| : : : : :
QY 359 AAGTIDGSGLNTGCTGEGDPFMDLRQDLCKPSE--IYEC-----400
| : : : : :
Db 256 AAGVI---SGLQVAGSGQAKYALDSKLVQDLARPEESALVAFGMNKTNDENYISY 312
| : : : : :
QY 401 -QKRP-----ILLHSGELTIPHPQOPDIVQIVTVGSLAIA-----438
| : : : : :
Db 313 DYOPTAPQOGEIANTVLFANGAL-IDGPPQPG-----NVGDDTLAAQIRQALDPKIK 364
| : : : : :
QY 439 -----IPGLTMSGRFREAIKKEFALY--GKMDTVIAGLSNVYHTTYTEEYO 489
| : : : : :
Db 365 AVILRVNSPGGSVAS-----ELIRAEALRAHNPVLSMGMAASGWTSTPANV- 418
| : : : : :
QY 490 AQRYEASTYIGPHTSATYQLF-----ROLAK-----AATPTVNMMS--SGDEPPE 535
| : : : : :
Db 419 -----IVASPTLGSIGIFGVINTFONSASIGVHTDGVATSPLDVSLTKALPE 470
| : : : : :
QY 536 FKNLILASLPN-----IADRAPICKHFGDVLQPAKPEYRGEVE--YIFGANKP--NSAEN 589
| : : : : :
Db 471 FSGMQMINTENGKTFIDLVATSRH-----KTPQVQDLAQNGVHWIGLAKSKGLVD 522
| : : : : :
QY 590 QTHQFTLYEKEKEDSVADQWIMVNDASWETRYMHK-----ILGLSNATYWHIP 640
| : : : : :
Db 523 Q-----LQDFDDAVKKAELAKLKTWQIMFVDEPSLDTLIGOMASVHAMLP 571
| : : : : :
RESULT 7
AD1129
probable peptidoglycan bound protein (LPXTG motif) lmo0435 [Imported] - Listeria mono
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AD1129
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloe
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Fslh,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tlerez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; M01D:21357279; PMID:11679669
A:Accession: AD1129
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2013 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CA098514.1; PID:q16409812; GSPDB:GN00177
A:Experimental source: strain Egd-e
C:Genetics:
A:Gene: lmo0435

Query Match 3.3%; Score 118.5; DB 2; Length 2013;
Best Local Similarity 19.8%; Pred. No. 8.5;

Matches 148; Conservative 88; Mismatches 266; Indels 247; Gaps 38;

44 FILADPOGSMNAF-----VSVELCMISORLRLKLEKSKYSIVR---RDNVILSAI 95
102 YLLSQGKYELEKEGDKVQAKLDLITKNDVEAKVERSNROLLSSISDKLFLCA- 160
96 HTHSGPAGFFQYTYLILASEGFSNRTFOYIVSGIMKSIDIAHTNLKPKIFINKGNVNV 155
161 --DSKATLANTYEQI-----TFNYSINFLDGS-----TNLKNKNIWIDERN-SNL 203
156 QINRSPSSYLLNPOSEARYSNTDKEMLVKLVLD--LNGE--DLGLI--SWFAIHPSVM 209
204 ELVYVPRKDTANRNRIKILTSYSGALTK--LFINLVNMISSGAPFDPIVYVAGYGAQGLP 261
210 NNSHFNPSDMNGAAY-----LFEQEKKGKYLPGGPFVAG-----246
262 MNLKATLSENGSSGTYTPSEKTTVTNLESSSNODY---SPITAGCNMAFNKELSY 317
247 -----FASNLGDV-SPNIGPHCVNTGE-----269
318 SLKRGYTIQWPELOKKSLEKSPKNLKLEVLKNGGVISVNTADPIVIRFGEFWSQL 377
270 SCQNDK-STCPNGSPSCMASPGQDMFESTHIIIGRIIYOKAKELVAS-----ASQEV 321
378 STVGKANVNLVNDDEKQVVEYGP-----INANIYOKIQVSMKIPADGVEGTEY 427
322 TGPVLAHOWNMVDVSQVLAHTVKTCKPALGYSPFAGTIIDG-VSGINITOGTGEDP 380
428 TGIYVNYDEDFITISIKI-----KAEVADSATSIADVSKVSTSI---SEGDI 472
381 F-WDLTRDQLLGKSEELVECOCKP-----LILHSGELTIHP 418
473 FEMGFMPVSSAAGVNDLEIVAPIBEGIKALSYIPNNNSMASIKLEIYONGKRVSMAP 532
419 -----WQPDIVDQIVVGSALAIAPG-----ELYTSGRRRERPAI 455
533 QTSSGMPFSKIDQSVNRLEKLTSLRSGIINDKMPYHTGIRMQNIGVAKGESF--TL 590
456 KKEPALKGMKDMT-VVIAGLSNVYTHYITTEVEYOQVQEAASITYGHTLSAYI-----509
591 RPESTITDPDKTSKTDITTTNSYGNVQVE-----KTSAPAKINGVFLSTAGTISGK 645
510 -----QLPFDLAKAIAITDVANMSSGPEPFFKNLIASLIPNADRAPIGKHFGVLOPAK 565
646 GFGSTIFPNDKINQSVRLSGYSGKLEMPY-----IFVVPKGIIDLETP-RNF--IQCP-- 696
566 PEYVGEVVEVIEVGAN-----PKNSAENQHTQFELTVEKYEEDSVALLQIMYNDASWET 619
697 --YR--STLNTYAPANGTNTLYPKRSSAD-----VKGKETLSGSTLY-----735
620 RFVYHKGILGSLNATYIWHIPDTAV-PGI 647
736 -----YWEAPDTGLAPGM 748

Db

RESULT 8

S70843

hemolytic protein hnda precursor - Haemophilus ducreyi

N:Alternate names: cytotoxin

C:Species: Haemophilus ducreyi

C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 08-Oct-1999

C:Accession: S70843

R:Palmer, K.L.; Munson Jr., R.S.

Mol. Microbiol. 18, 821-830, 1995

A:Title: Cloning and characterization of the genes encoding the haemolysin of Haemophilus

A:Reference number: S70843; MUID:96422469; PMID:8825086

A:Accession: S70843

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1175 <PAL>

A:Cross-references: EMBL:U32175; NID:9151070; PIDN:AMC43538.1; PID:9151072

C:Genetics:

A:Gene: hnda

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-1175/Product: hemolytic protein hnda #status predicted <MAT>

Query Match 3.2%; Score 113.5; DB 2; Length 1175;

Best Local Similarity 15.5%; Pred. No. 8.5;

Matches 103; Conservative 120; Mismatches 252; Indels 189; Gaps 22;

8 VGRDCTGOVDINLMGCKGNQARGL-LRLRSRAITLDP-----DGSNR 54
116 LKQETLAKIADYIIV--NPMGMSDGGCFINISASLVGNPNTDGLVLRIDGSHN 173
55 MAFVSELMISORLRLKLEKSKYSIVRQVILSAHTHSGPAGFFQYTYLILAS 114
174 ---ISTTDIAEKTNLILAPVYVNIKINIGKOVNIIITHNEIN-VNNDQLSISVLPN 229
115 EG-----FSNRIFQYVSGIMK---SIDAHTNLKPKIFINKGN 151
230 KGKVLGDKIAGSIQANRIRIHTDNRATLIDETGLKGTNSVYIGAGMLKY-----HGH 283
152 VANOQINRSPSSYLLNPOSEARYSNTDKEM-----LVKLKLVLDLNGEDLGLI-----199
284 IDTKNNNNKNNKYYLAKSTQEQVSMITGDDINLNTQLTITNANLTKGNIVLKAASQ 343
200 -----SWFAIHPSVMNNSHFNVS---DNMGYAAYLFEQEKNGKYL 237
344 FGTERTLNKHYEKNRRSSGSWYHNNETSKEEINVTISIKGDNVGIVA-----DQCAI 396
238 PGQGFVAGFASNSLIGVSPNIGPHCVNTGESCNDKSTCPNGSPSCMASPGQDMFE 297
397 DGOSLKITAQNTTTLGKQGVTLRGTONINREKEKFEFNETSN-----439
298 STHIIIGRIIYOKAKELVASQVGVPLAHOWNMVDVSQVLAHTVKTCKPALGYS 357
440 --LTTGKFFKNQOSQYIASELDIKGNLKITGGGDIKGVVYRTGGDELVA-----N 489
358 FAAGTIDVSGINITOGTGEDPFWDLRLDQLGKSEELVECO-----401
490 TGSTVNAENILN--SNINDYENWVGIGGSNTYVKKNSITLQOCADLVINGKTYIDSEK 548
402 -----KRPKILHSGELTIHPWQPDIVDQIVVGSALAIAPGELTITNSGR 449
549 GVKISGRSLVSGKEALVRAHKGRLVIDV--KNEITSLVIRKGVV-----DITKASAK 601
450 RPRE-----AIKKEPFL-----YGMKDMYVYAGLSNYTHYITTE 486
602 DFKONSVRGSTVSESNNLKLVSADGIEVAGSLYKSLGDLVTKDCKSNLYVKGQONIM 661
487 EYQQRVEASTIYGPHTLSAYIQLFRDLAKAIAITDVANMSSGPEPFFKNLIASLIPN 546
662 ERESEK--SGLTGK-----NLDPSLIGIDYTKNIAVSVTGEVLPDLMQKIDN 707
547 IADR 550
708 LQEK 711

Db

RESULT 9

B81989

hypothetical protein NMA0688 [imported] - Neisseria meningitidis (strain Z2491 serogr

C:Species: Neisseria meningitidis

C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: B81989

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo

R.; Holtroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491

A:Reference number: A81775; MUID:20222556; PMID:10761919

A:Accession: B81989

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2015 <PAR>

A:Cross-references: GB:AL162753; GB:AL157959; NID:97379120; PIDN:CAB83974.1; PID:97377

A:Experimental source: serogroup A, strain Z2491

C:Genetics:
A:Gene: NMA0688

Query Match
Best Local Similarity 18.9%; Pred. No. 22;
Matches 111; Conservative 86; Mismatches 204; Indels 272; Gaps 32;

3.2%; Score 113; DB 2; Length 2015;
109 LYLASEGFSNRTFOYIVSGIMKSIDIAHTNLK-PGKIFLNKGNVANOINRSPSYLLN 167
111 LALLADNITAKT-----TNLNTPCNLVHTGKDLNLVADKDLASAIIH 554
168 POSERARISNTDKEMLVAKLVNDGEDGLISFWFAIHVPVSMNSNHFVSDMGR----- 223
555 LKSDNAHITGTSTLTASK--DM-GVEAGLL-----NVTNLTNRNSGLHIOAK 603
224 -----AAVLEOEKNGKYLPGQG-----PFVAGFASNLGDVSPNIGPHC- 264
604 GNIOGRLNTKLNMAKALETTALGCGNIYSDGLHVASDGHVSLANGNADFTGHHTLAKAD 663
265 VNTGE-----SCDNDKSTCPNGSPCMASGPGDM----- 295
664 VVAGSVKGRKLKADNTNITSSSG--DITVAGNGIQLGDKGRNSINGKHISIKNGGNA 721
296 -----FESTH-----II 302
722 DLKNLNVHAKSGALNHSRALSTENTKLESTHNTLNQHERVTLNQDVAHRLST 781
303 GRITVOKAKELYASAOEVTGPVLAHQVNMVDVSNLTATHTVTKCPALGYSFAGT 362
782 GSOIQWQNDK--LPSANKVLAVNGVLA-----LNARYS-----QIADNT 816
363 IDGVSGNLITOGT---EGDPFWDLRDQLGKPSSEIVECQPKPILH---SGELTI 415
817 TLRAGAINLTAGTALVKRGNINMSTVSTKLEEDNAB-----LKPGLRINIEAGSGLTI 871
416 -PHRQPDIVDQIVTVGSLAIAIRG-----ELTMSGRFRRAIKKEFLYGM--- 464
872 EPANRISANTDLSIKTGGKLLLSAKGNAGAPSAOVSLKAKNIRLVGETDLRKSKIT 931
465 --KDMTVV-----TAGLSNVYTHYITVE-----EYQARYEASTIYGPH 504
932 AGKNLVVATTGKLNIEAVNNSPVSFPQKAELNQSKHELDQIAQLKKS-----PKS 987
505 -----LSAIOLEPRDLAKA-----IATDVVAMS-----SGPEPP 534
988 KLIPTLQOEERDLAFYIOAINKEKPKRGEYLOAKLSAQNIDISAGLEISGSDIT 1047
535 FPKNL-----IASLIPNIAADRAPIGK-----HFGDVL--QPAKPEYRVG 571
1048 ASKKLNLHAAGVLPKRADEMAAILDGIITDOYEIETKPKYKSHYDKAALNKPRLTGRIG 1107
572 EVVEV-----IFVGAN-----PKNSAENQTH 592
1108 VSIHAAALDDARIITIGASEIKAPSGSIDIKAH 1140

RESULT 10

S75251
hypothetical protein slr1028 - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: S75251

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimoto, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

S.
A:Reference number: S74322; MUID:97061201; PMID:8905331

A:Accession: S75251

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-3972 <KAN>

A:Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PIDN:BAAL17165.1; PID:g165
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: *Synechocystis* hypothetical protein slr1028

Query Match
Best Local Similarity 21.1%; Pred. No. 78;
Matches 80; Conservative 60; Mismatches 130; Indels 110; Gaps 21;

3.1%; Score 112; DB 2; Length 3972;
105 FOYTLVILASEGFSNRTFOYIVSGIMKSIDIAHTNLKPKI---FTNKN-----VANY 155
1645 FSYPI-----QGTENLFLIDPG-----SSNTTIGPVEVTVYFENNDRTPFESVAPV 1693
156 QINRSPSYLLNQSERARISNTDKEMLVAKLVNDGEDGLISFWFAIHVPVSMNSNHF 215
1694 QINMA-----TMGDDQTVTGTATVTEANDSLALID-----SGFI 1728
216 VNSDN--MGYA--AYLEOE-----KNGKYLPGGQPFVAG-----FASNL-----GD 254
1729 INSDNPAIGVILASAFNSDGTLAYVAVGNRGISDSQGNVNLNCTVOILFSCDILSGSGS 1788
255 VSPNLT-----GPHCVNTGESCDNDKSTCPNGSPCMASGPGQDMFESTHITIGRIYOK 309
1789 LSTTILNGNDPGLVLTINIQADGNQRL-----SLATAGD-----IDGDSIPDLVIGAP 1838
310 AKELYASASOEVTPVLAH--QVNMVDVSNLTATHTVTKCPALGYSFAAGTIDGVS 367
1839 MNGNFAGAVYVIGSYLSNOKGQIIDVTNLSTKPN-----TMGEFVNGNEAEDLA 1888
368 GUNITOGTTEGDFPWLTRDQLGKPS-----EIVECQKP---KPIILHSGE- 412
1889 GFSVYVGNFDGDSYGDIVGAPYAKDSNGNRGVYLVAGFAGSAPDSISPTVITSGKS 1948
413 LTRHPQPDIVDQIVTVG 432
1949 FDIPIPD-QPN-PPSQTLTVG 1966

RESULT 11

T43369

heat-shock protein HSP60 precursor, mitochondrial - *fission yeast* (*Schizosaccharomyces*

C:Species: *Schizosaccharomyces pombe*

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000

C:Accession: T43369

R:Toshida, H.; Yanagi, H.; Yura, T.

Gene 167, 163-166, 1995

A:Title: Cloning and characterization of the mitochondrial HSP60-encoding gene of *Sch*

A:Reference number: 222456; MUID:96144268; PMID:8566770

A:Accession: T43369

A:Status: preliminary; translated from GB/EMBL/DDDB

A:Molecule type: DNA

A:Residues: 1-582 <TOS>

A:Cross-references: EMBL:D50609; NID:g1229150; PIDN:BAAL09171.1; PID:g1229151

A:Experimental source: haplotype h-1eul-; strain Hml23

C:Genetics:

A:Gene: mcp60

C:Function:

A:Description: essential for cell viability at all temperatures

C:Superfamily: chaperonin groEL

C:Keywords: mitochondrion

Query Match
Best Local Similarity 23.8%; Pred. No. 5.5;
Matches 63; Conservative 43; Mismatches 92; Indels 67; Gaps 14;

133 IDIAHTNLKRGKIFINKGNVANOINRSPSYL-----LNQSERAR-----YSSN 178
336 IDVSIKADPHL-----GSGSTVTVKEDITIMKKGAGDHAKVVDREQIRGVADPNLT 391
179 TDKEMVLKLVNDGEDGLISFWFAIHVPVSM--NSNHFVNSDNMGYAAVLEOEKNGKY 237
392 SEKEKLEERLAKLSG--GIAVIRKVASSEVEVNEKKRIRYDALNAVAAV-----SEGYL 444
238 PGCG-PFVAGFASNLGDVSPN-----ILPHCVNTGESCDNDKSTCPNGSPCMASGPG 292

Db 445 PGAGTSEVVK--ASLRIDIPNNPDOKLGEIVKRA-----ITRPA 483
Oy 293 ODMESTHIIIGRIYOKAKELYASASQEVTPVLAHOWVMEDVSOVLAHTHVKKCKP 352
Db 484 OTIENMGLECNLVGLKLELYKEFN--IGYDIKDFVDLNEIGV-IDPLKVVRT-- 537
Oy 353 ALGYSFAAGTIDVSGVSGINTGCTTE 377
Db 538 -----GLVD-ASGVASLMTGTE 553
RESULT 12
C2HU
complement C2 precursor [validated] - human
N:Contins: classical-complement-pathway C3/C5 convertase (EC 3.4.21.43) C2a subunit; cl
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1988 #sequence-revision 30-Jun-1988 #text-change 23-Mar-2001
C:Accession: A25971; J50281; A25290; A05289; A37540; A26506; B26506; I56179; I54
R:Wu, L.; Morley, B.J.; Campbell, R.D.
Cell 48, 331-342, 1987
A:Title: Cell-specific expression of the human complement protein factor B gene: evidenc
A:Reference number: A25971; MUID:87102880; PMID:3643061
A:Accession: A25971
A:Molecule type: DNA
A:Residues: 694-752 <MUL>
A:Cross-references: GB:M15082; NID:g187699; PIDN:AAA59624.1; FID:g467309
R:Horluchi, T.; Maccon, K.J.; Kidd, V.J.; Volanakis, J.E.
J. Immunol. 142, 2105-2111, 1989
A:Title: cDNA cloning and expression of human complement component C2.
A:Reference number: J50281; MUID:89156483; PMID:2493504
A:Accession: J50281
A:Molecule type: mRNA
A:Residues: 1-532, 'F', 534-752 <HOR>
A:Cross-references: GB:M26301
R:Bentley, D.R.
Biochem. J. 239, 339-345, 1986
A:Title: Primary structure of human complement component C2.
A:Reference number: A25290; MUID:87127920; PMID:2949737
A:Accession: A25290
A:Molecule type: mRNA
A:Residues: 1-752 <BEN>
A:Cross-references: GB:X04481; NID:g34627; PIDN:CAA28169.1; FID:g34628
R:Bentley, D.R.; Porter, R.R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1212-1215, 1984
A:Title: Isolation of cDNA clones for human complement component C2.
A:Reference number: A05289; MUID:84144868; PMID:6199794
A:Accession: A05289
A:Molecule type: mRNA
A:Residues: 588-717 <BE2>
A:Cross-references: GB:K01236
R:Gagnon, J.
Philos. Trans. R. Soc. Lond. B Biol. Sci. 306, 301-309, 1984
A:Title: Structure and activation of complement components C2 and factor B.
A:Reference number: A37539; MUID:85038851; PMID:6149575
A:Accession: A37539
A:Molecule type: protein
A:Residues: 137-149, 'AG', 150-171, 454-466, 574-717 <GAG>
A:Note: two glycosylation sites were determined
R:Parke, C.; Gagnon, J.; Kerr, M.A.
Biochem. J. 213, 201-209, 1983
A:Title: The reaction of iodine and thiol-bloking reagents with human complement compo
hioi group.
A:Reference number: A37540; MUID:83308518; PMID:6555044
A:Accession: A37540
A:Molecule type: protein
A:Residues: 244-267, 'G', 269 <PAR>
R:Kerr, M.A.; Gagnon, J.
Biochem. J. 205, 59-67, 1982
A:Title: The purification and properties of the second component of guinea-pig compleme
A:Reference number: A26506; MUID:83022288; PMID:6922702
A:Accession: A26506
A:Molecule type: protein

A:Residues: 244-248, 'SX', 251, 'XK', 254-256 <KER>
A:Accession: B26506
A:Molecule type: protein
A:Residues: 21-23, 'X', 25-28, 'XUX', 32-33, 'S', 35-38, 'X', 40, 'X', 42-44, 'X', 46 <KE2>
R:Shih, Y.; Zhu, Z.B.; Maccon, K.J.; Volanakis, J.E.
J. Immunol. 151, 170-174, 1993
A:Title: Structure of the human C2 gene.
A:Reference number: 156179; MUID:93315833; PMID:8326124
A:Accession: 156179
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-532, 'F', 534-752 <RES>
A:Cross-references: GB:I09708; NID:g179663; PIDN:AAB97607.1; PID:g298124
R:Bentley, D.R.; Campbell, R.D.; Cross, S.J.
Immunogenetics 22, 377-390, 1985
A:Title: DNA polymorphism of the C2 locus
A:Reference number: 154419; MUID:86032058; PMID:2997031
A:Accession: 154419
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 21-46 <RE2>
A:Cross-references: GB:M15549; NID:g187764; PIDN:AAA59649.1; PID:g187765
C:Genetics:
A:Gene: GDB:C2
A:Cross-references: GDB:119731; OMIM:217000
A:Map position: 6p21.3-6p21.3
A:Introns: 16/1; 86/1; 148/1; 206/1; 239/1; 283/3; 330/1; 377/1; 407/1; 454/1; 485/3;
C:Complex: The proenzyme forms a complex with C4a and is activated by cleavage into C
C:Function:
A:Description: cleaves complement C3 and complement C5 alpha chains
A:Pathway: complement classical pathway
C:Superfamily: complement C2; complement factor H repeat homology; trypsin homology;
C:Keywords: complement classical pathway; duplication; glycoprotein; hydrolase; plasm
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-243/Product: complement C2b subunit #status experimental <C2B>
F:89-144/Domain: complement factor H repeat homology <FH1>
F:151-204/Domain: complement factor H repeat homology <FH2>
F:244-752/Product: complement C2a subunit #status experimental <C2A>
F:252-442/Domain: von Willebrand factor type A repeat homology <VFA>
F:471-739/Domain: trypsin homology #status atypical <TRY>
F:24-64, 51-84, 89-131, 117-144, 151-191, 177-204, 463-581, 492-508, 584-600, 638-665, 675-705/
F:29, 112, 290, 333, 467, 471/Binding site: carbohydrate (Asn) (covalent) #status predicte
F:243-244/Cleavage site: Arg-Lys (complement subcomponent C1s) #status experimental
F:507, 561, 679/Active site: His, Asp, Ser #status predicted
F:621, 651/Binding site: carbohydrate (Asn) (covalent) #status experimental
Query Match 3.0%; Score 109; DB 1; Length 752;
Best Local Similarity 21.6%; Pred. No. 9;
Matches 93; Conservative 58; Mismatches 151; Indels 128; Gaps 23;
Oy 262 PHCVNTGES--CDNDKSTCPNGSGMCMASPGQDMESTHIIIRIYOKAKELY--ASA 317
Db 133 PNGMDEDTAVCDAGACHCPNGISL---GAVTGRFGH-GDKVRRSSNLVLGSS 187
Oy 318 SOEVTGPIVLAHOWNMNVDVSOVLAHTHVYTCRPAFGYSPAAGTIDVSGVSGINTGCTTE 377
Db 188 ERECGNGV---WSGEPICRQPYSDPEDVAPALGTSRS---HMLGATNPTOTKTE 239
Oy 378 GDPFMDTLRDQLGKPSR-----EIVECQKPRP---ILLHSGELTTPHPQPD 422
Db 240 S-----LGRKIQIORSGLNLYLLDCQSVSENDLFIFKESASIMVDRIPSFE 288
Oy 423 I-VDQIVTVGSLAIAIPGELTITMSGRPRFRAIKKFPALXGMQDMVTVAGLSNVYTHY 481
Db 289 INVSVAIT---FASPEKVLMSVLNDSR-----DMTEVLSLEN----- 325
Oy 482 ITTYEYO---AORYEASTIY-----GPHLSAVIQLFRLAKAI--ATDTVA 525
Db 326 -ANKHDHNGTGTYTAAALNSVYLMNNQKRLGMEYMA-----WQIRRAIILLTGKS 379
Oy 526 NMSSGPPEPFKRLIASLIPNIADRADIGKHFQGVLOAPAKPEYRVG-----EVEYV 576

Db 380 NMGSGPK-----TAVDHIREFILINOKRNDYD----IYAIVGKLDVDRELNEL 426
Qy 577 IFVGANPKNSAENQ-----HQT-----LTVKEEDSVADMOIYNADSMETRFYWH 624
Db 427 -----GSKKDEGHAFILQDTKALHOFHEHMLDVSKLFTDTICGVMSANASQOEKTPMH 481
Qy 625 KGLGLSNMT 634
Db 482 VTIKPKSOET 491

RESULT 13
T05174
Hypothetical protein T6K22.50 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05174
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirke, W.; Stiekema, W.; Bancroft, I.; Men
submitted to the Protein Sequence Database, August 1998
A:Reference number: Z15400
A:Accession: T05174
A:Molecule type: DNA
A:Residues: 1-1736 <BEV>
A:Cross-references: EMBL:AL031187
A:Experimental source: cultivar Columbia; BAC clone T6K22
C:Genetics:
A:Map position: 4
A:Intons: 42/3; 106/3; 176/1; 304/1; 366/3; 397/2; 430/1; 484/1; 704/3; 799/2; 837/3;
A:Note: T6K22.50

Query Match 3.0%; Score 109; DB 2; Length 1736;
Best Local Similarity 20.7%; Pred. No. 34;
Matches 111; Conservative 67; Mismatches 184; Indels 174; Gaps 29;

Qy 227 LFEQEKNGKYLPGQ-----GPFVA-----GFAS-----NLGDVSPNIIAP 262
Db 694 LITLENKKYTL-GQSLYTPREISFTYVICTGHSNNDQITKGVIMHFSKGPYRP--LTP 750
Qy 263 HCV-----NTGESCONDKSTCPNGSPSCMASGPGDMFESTH----- 300
Db 751 DVQKNGIGILYVRNPGDS-----RVCEVNFPCILDMVEGSELTYIQTSSMKIKIS 806
Qy 301 ----IIGRIYOKAKELYA-----SASQEVTPRYLAHQVNMMD--VSQVLNATHYKTKC 351
Db 807 PKTTIIGESVASVAKSSARGPSFSPALIKPDIAAPGLTLTPRIPTDEDRREFYVS-- 864
Qy 352 PALGVSFAAGTIDGVSGNLITQGTTEGDFPMDTLRDQLLGKPESEIVECQPKPILLHSG 411
Db 865 ----GTSMATPVLAGIAL----- 879
Qy 412 ELTIPHP-WQPDIVDQIVTVGSLAIAIP-GEIITMSGRFRREAIKKEPA-----LYGN 464.
Db 880 -LKISHPNMSPAVIKSALVTT--AMKTPYGERLTVDCGNKYVADAIFYGGGLVNLKRA 935
Qy 465 KDMTVIAGLSNVYTHYITTEEYQAKREASTIYG---PHTLSAYIQL-----FRDL 515
Db 936 TPGGLYDMIDINDYTHLCOSQLTYDKKVSALGNVNNKCPSSSSSILDLNPSITIPDL 995
Qy 516 AKAI-ATDVANMSSGPEPPFFKNLIALIPNIDRAPIGKHEGDVLAQAKPEY-----R 569
Db 996 KGTIVNTRTYTNV--GRVKSYPKPYI-----EAPFG--FNVSYSKKKLKFNKTRKK 1042
Qy 570 VEEVEEVITVGANPKNSAE--NOTHOTFLTVEKEDSVADMOIYNADSMETRFYWHKGI 627
Db 1043 LAFTIYVHILGVRHDSSELVESHQRMLESFESAEARESIYVN-----YHNGF 1093
Qy 628 LGLSNMTIYWHIPDPA-----YPCGIYRIYFGHNKOKELKRAVLAEIGISSPF 677
Db 1094 SGFPA-----RLTDSQAKQLSDRPDVFV--APNRKVE-LOSTRILYDLGLSPSF 1140

RESULT 14
B81192

hemagglutinin/hemolysin-related protein NMB0497 [imported] - Neisseria meningitidis (C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: B81192
R:Retlein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. et al.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizza, M. Science 287, 1809-1815, 2000
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappaport, R. A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: AB10000; MIMD:20175755; PMID:10710307
A:Accession: B81192
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1975 <TEF>
A:Cross-references: GB:AE002406; GB:AE002098; NID:97225720; PIDN:AAF40929.1; PID:9722
C:Genetics:
A:Experimental source: serogroup B, strain MC58
A:Gene: NMB0497

Query Match 3.0%; Score 109; DB 2; Length 1975;
Best Local Similarity 18.6%; Pred. No. 42;
Matches 129; Conservative 86; Mismatches 206; Indels 272; Gaps 31;

Qy 109 LYLASEGFSNRTFOYIVGIMKSIDIAHTNLK-PEKIFINGVANVAOINSPSYLLN 167
Db 508 LALLADNDITAKT-----TNLTNPENLVHVGKDLNVDLDSASIH 551
Qy 168 POSERARYSSNTDKEMLVKLVDNCEGLISWFAIHVSMNSNHFNSDMGW----- 223
Db 552 LKSDNAHITGTSKITASK-----DMGV-----EAGSLNVTNLTNRNSGHLHQAOK 600
Qy 224 -----AAYLEQEKNGKYLPGQ-----PFVAGFASNNLGDVSPNIIAPHC- 264
Db 601 GNILRNTKLNAKALETALQGNIVSDGLHVASADGHVSLLANGNADEFTGHNTLAKAD 660
Qy 265 VNTGE-----SCDNDKSTCPNGSPSCMASGPGDM----- 295
Db 661 VNAGSVGKRLKADNTNITSSG--DITLVAGSIGLDGKORNSINGKHSIKNGNA 718
Qy 296 -----FESTH----- 718
Db 719 DLKLNVAHKKSGALNHSRALSIENTKLESTHNTLNQHERVTINQDAYAHRHLSIT 778
Qy 303 GRIIYOKANELYASASQEVTPRYLAHQVNMMDVSQVLNATHYKTKCPALGYSFAGT 362
Db 779 GSDIQWQNDK--LPSANKLVANGVLA-----LNARYS-----QTADNT 813
Qy 363 IGVSGSLNITQGT-----EGDFPMDTLRDQLLGKPESEIVECQPKPILLH-----SGELTI 415
Db 814 TLKAGALNLTAGTALVKRGNINNVSTVSTKLEDNAL-----LKPPLAGRLNIEGSGTLTI 868
Qy 416 -PHPQPDIVDQIVTVGSLAIAIPG-----ELTWSGRFRREAIKKEPALYGM----- 464
Db 869 EPANRISATHTDLSIKTGKLLLSAKGNGAPSAQVSSLEAGNIRLVGELDLRSKIT 928
Qy 465 --KDMTV-----IAGLSNVYTHYITTE-----EYQAKREASTIYGPHF 504
Db 929 AGKNLVAVATKGLNLEAVNNSFSNFPQKAELNOKSKELEQOIAQLKKS-----PKS 984
Qy 505 -----LSATIQLFRLAKA-----IATDVANMS-----SGPEP 534
Db 985 KLIPLOEBRDLAFYIQAINEVKKKKRKEVEYQAKLSAQNIDILISAGIEISSDIT 1044
Qy 535 FFKNL-----IASLIPNIDRAPIGK-----HFGDVL--OPAKPEYRVG 571
Db 1045 ASKKNLNLHAGVLPKRADEBAAILLDGTTDQYEIKTPPYKSHYDAALNKRSLRGLRG 1104
Qy 572 EYEV-----IFVGAN-----PKNSAENQTH 592
Db 1105 VSIHAAALDARITIGASEIKAPSGSIDIKAH 1137

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 12:46:00 ; Search time 55 Seconds
(without alignments)
1425.912 Million cell updates/sec

Title: US-09-937-521-14

Perfect score: 3582
Sequence: 1 FSGYIGVGADCTGVSDI.....KPAVILAFEGISSPEVYTT 682

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 11492915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCR_NEM_PUB pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEM_PUB pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCRUS_PUBCOMB pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEM_PUB pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1090.5	30.4	646	10 US-09-777-710A-1	Sequence 1, Appli
2	1090.5	30.4	670	10 US-09-777-710A-15	Sequence 15, Appl
3	879	24.5	280	10 US-09-739-907-146	Sequence 146, App
4	874	24.4	318	10 US-09-739-907-76	Sequence 76, Appl
5	131.5	3.7	1565	9 US-10-242-056-59	Sequence 59, Appl
6	127	3.5	27	10 US-09-739-907-150	Sequence 150, App
7	125	3.5	26	10 US-09-739-907-148	Sequence 148, App
8	116.5	3.3	766	10 US-09-815-242-10572	Sequence 10572, A
9	113	3.2	1981	9 US-09-928-457-38	Sequence 38, Appl
10	113	3.2	2015	9 US-10-066-551-1	Sequence 1, Appli
11	111.5	3.1	627	10 US-09-338-723A-4	Sequence 4, Appli
12	111.5	3.1	627	12 US-10-080-210-4	Sequence 4, Appli
13	106	3.0	1376	9 US-10-303-685-14	Sequence 14, Appl
14	106	3.0	1406	9 US-10-303-685-15	Sequence 15, Appl
15	104	2.9	878	9 US-10-165-049-2	Sequence 2, Appli
16	104	2.9	878	9 US-10-165-049-3	Sequence 3, Appli
17	104	2.9	878	10 US-09-905-983-48	Sequence 48, Appl
18	101.5	2.8	929	9 US-10-156-761-14135	Sequence 14135, A
19	101	2.8	929	10 US-09-815-242-12903	Sequence 12903, A

20	101	2.8	1399	10 US-09-815-242-5179	Sequence 5179, Ap
21	100	2.8	30	10 US-09-739-907-153	Sequence 153, App
22	100	2.8	613	10 US-09-996-620-10	Sequence 10, Appl
23	100	2.8	621	10 US-09-996-620-18	Sequence 18, Appl
24	100	2.8	1021	10 US-09-815-242-5471	Sequence 5471, A
25	100	2.8	1021	10 US-09-815-242-12544	Sequence 12544, A
26	99	2.8	1356	9 US-09-757-781-2	Sequence 2, Appli
27	98	2.7	24	10 US-09-739-907-152	Sequence 152, App
28	97.5	2.7	882	9 US-10-097-340-37	Sequence 37, Appl
29	97	2.7	527	9 US-10-128-714-3378	Sequence 3378, Ap
30	97	2.7	556	9 US-10-128-714-3561	Sequence 3561, Ap
31	97	2.7	556	9 US-10-128-714-8378	Sequence 8378, Ap
32	97	2.7	556	9 US-10-128-714-8561	Sequence 8561, Ap
33	97	2.7	1115	9 US-10-205-823-160	Sequence 160, App
34	97	2.7	1115	9 US-09-781-558-2	Sequence 2, Appli
35	96.5	2.7	439	10 US-09-374-586-2	Sequence 2, Appli
36	96.5	2.7	454	10 US-09-835-147-6	Sequence 6, Appli
37	96.5	2.7	464	10 US-09-835-147-30	Sequence 30, Appl
38	96.5	2.7	464	10 US-09-835-147-27	Sequence 27, Appl
39	96.5	2.7	473	10 US-09-835-147-29	Sequence 29, Appl
40	96.5	2.7	474	10 US-09-835-147-28	Sequence 28, Appl
41	96.5	2.7	476	10 US-09-835-147-3	Sequence 3, Appli
42	96.5	2.7	476	10 US-09-835-147-4	Sequence 4, Appli
43	96.5	2.7	478	10 US-09-835-147-8	Sequence 8, Appli
44	96.5	2.7	487	10 US-09-835-147-26	Sequence 26, Appl
45	96.5	2.7	502	9 US-10-092-063-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1

US-09-777-710A-1

Sequence 1, Application US/09777710A

Patent No. US20020058305A1

GENERAL INFORMATION:

APPLICANT: OKINO, No. US20020058305A1om et al.

TITLE OF INVENTION: CERAMIDASE GENE

FILE REFERENCE: 1422-0458P

CURRENT APPLICATION NUMBER: US/09/777,710A

CURRENT FILING DATE: 2001-02-07

NUMBER OF SEQ ID NOS: 18

SOFTWARE: PatentIn version 3.0

SEQ ID NO 1

LENGTH: 646

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-777-710A-1

Query Match 30.4%; Score 1090.5; DB 10; Length 646;

Best Local Similarity 35.9%; Pred. No. 9e-87; Indels 75; Gaps 14;

Matches 250; Conservative 118; Mismatches 254;

QY	4	Y Y I G V R A D C T G V S D I N L M G Y G K N G Q N A R G L T R L F S R A F L A D P D S G N R A F V S E L C	63
DB	5	Y F G L G K A D I T G E A N A V G M G Y S L E Q T A G I H M Q M A R A F I E D A S G R L V Y V N T D L G	64
QY	64	M I S Q R L E V L K R L E S K Y G S L Y R D N V I S A I H T S G A G F O Y L Y I L A S G F S N R F Q	123
DB	65	M F F Q A V H L K V L A R L K A K Y G Y D E N N V M L A A H T H S G G G F S H Y A M Y N L S V L G F O E K T F N	124
QY	124	Y V S G I M K S I D I A H N L K P G K F I N K G V A N Q I R S P S S Y L N Q S E R A R S S T D E M	183
DB	125	A I V D I V R S I E A Q A R L O P G R L F Y S G E L R N S R N K S L S L K N P - D I A G E D C I D P M	182
QY	184	I V L K L V D I N G E D I G L I S W F A I H P V S M N N S N H F V S D N G M Y A Y L F E O E K N K Y L P G O G P F	243
DB	183	S Y L S V D A N G E L A G I A S M P H V S T S M T A N H L I S P D N K G Y A S Y H M E H V S R - - - - K S G F	237
QY	244	V A G F A S S N I G D V S P I L G P H C V N T G E S C D N K S T C P N G S P S M C M A S G G Q D M F E S T H I L G	303
DB	238	V A A F A Q T V A N G N L S P M L - - - - - N L K P G S G P F D E F D N T R E I G	273

```

QY      30  RIYKAKETELVYASQSEVTPVLAHVNMMDVSVOLNAT--HTVKCKRPALGYSPAG 361
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      274 LRQPAKAYELAGAOGEVGLGELDSRFREVDFTRLRIREFEFDGQPRQLCTAIGTSLAG 333
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      362 -TIDVSGSLNTIGTGGDFPMDLRLDOLGKPSSEIVECQPKXILHSGELTIPHPMQ 420
      S E D G G P I G L E G ---NNPFLSALIGLITGVPPGELVQCAEKTLIADTGNNK-PPWT 389
Db      334 STEDGPGPIGLSEG---NNPFLSALIGLITGVPPGELVQCAEKTLIADTGNNK-PPWT 389
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      421 PDIVDQIVTVGSLAIAIPGELITTMSCGRFEAIKKKEPALYGMKMDTVIAGISNYTH 480
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      390 PVALPIOMFPIGGLLELGLGAPAEFTYWAGIRIRAVQAASAAGIRH--VWNGANAYAS 447
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      481 YITVEYQAOQRLEASTIYCPHTLSAYIQLEFRDLAK-----IATDTVA-----N 528
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      448 YVTTREYIAOEEGGSTLGPHTQAAIQQLLEVDAVLAIRRELPVETSALNPDISCCOMN 507
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      527 MSSGPEPPEFKNLIALENIADRAPICGHEGSDVLOAPKRYRGEVEVEYLFVGANFNS 586
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      508 FQGVG-----VADDPYIGSFEDVLDQPRRESRIQDKVTAVFYGHRRN- 551
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      587 AENQVQOTELTYEKT-EDSVADMQIMYNDASKEFTFYHHKGIILGSLNATYIWIHPDAYP 645
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      552 -DLRTEKTELEVNNICKDQKQTPVAVADNDMDIDTQYRWRERGISASKATISWISIPCTEP 610
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      646 GIVRIYFEGNKKQELLKPAVILAIEAGTSSPEVYTT 682
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      611 GHYIIRHIGNAKNFTQK---ISELGGSTRRSEVLTCT 644

```

RESULT 2

```

US-09-777-710A-15
: Sequence 15, Application US/09777710A
: Patent No. US20020058305A1
: GENERAL INFORMATION:
: APPLICANT: OKINO, NO. US20020058305A1ommu et al.
: TITLE OF INVENTION: CERAMIDASE GENE
: FILE REFERENCE: 1422-0458P
: CURRENT APPLICATION NUMBER: US/09/777,710A
: CURRENT FILING DATE: 2001-02-07
: NUMBER OF SEQ. ID NOS: 18
: SOFTWARE: PatentIn version 3.0
: SEQ. ID NO 15
: LENGTH: 670
: TYPE: PR1
: ORGANISM: Pseudomonas aeruginosa
US-09-777-710A-15

```

Query Match	Score	DB	Length
30.48;	1090.5;	10;	670;

Matches 250; Conservative 118; Mismatches 254; Indels 75; Gaps 14;

```

OY      4 YUUGVRACTGQVSDINLMGYGKGNQVNRAGILTFEFRAITLDDPQDSNMMAVSELC 63
Db      29 YRGLCKAIDITGEAAEVMGMYSSLSLEOKTAGIHNRMQARATIEEASGRRLYVNTDLG 88
OY      64 MISQRLRELYLRLESKYGSILYRDNVILSAIHTHSGPAGEFOYTYLLASSEGSNFTQ 123
Db      89 MFEQVAHLVLYLRKAKYPCGYVDENNVNMLAATHHSGSGGESHYAMVLYLVLGPOKFTFN 148
OY      124 YIYSSGIMKSIDIAHNTLNKPKCFEIKKGVANVQINRSPSYLLNPOSPRARYSNTPKEM 183
Db      149 AIYDGIYVRSIERQARLOPGRLEFYSGGELIRNASRRSLSLSHLKNP--DIAGYEDGIDPOM 206
OY      184 LVVLKLDLNGEDGLISFPAIIPRVSMNNSNHPVNSDNGGAAUYLFEEQKXNGYLPQGGPF 243
Db      207 SYLSPFDANGELAGALSMFPRVHSTSMYANHLISPDNKGYSATYMEHDVSR-----KSGF 261
OY      244 VAGFASSNLGDVSPNLTGPHCVNTEGSCDNDKSTCPNGSPMCMAASGPGQDMFEHTIIG 303
Db      262 VAAFAQTNGNSPNL-----NLKPGSGPFDNEFDNTRBIG 297
OY      304 RIYQAKKELLYASASQEVYTGVLANKQVNNMTDVSVDLNLAT--HTVATKCPALDYSRAAG 361

```

```

Db      298 LRQFAKAYELAGAOEVLGELDSNRFVDFRLPIREFIDGQPRDLCTAALGTSLAG 357
Qy      362 -TIDVSGLNTIOGTEDGPEWMDRLDOLLGKPESEIVECOKPRILHSGELTIPIHWQ 420
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      358 STEDGPGPLGLEGG---NNPFLSALGGLTGPBPDELQCOAEKTIILADGNKK-PPYMT 413
Qy      421 PDIVDVOITVGSILAIHSGELTMSGRRPEAKIKKFKALYSGKMDTIVTLAGSNYTH 480
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      414 PVLPLPIOMERIGOLELLGAPAEFTYAGARRIRRAQAASEAGLRH--VFNNGATAYAS 471
Qy      481 YITVEEYOARENAASTYTPGHTLSAYITLERDLKA-----IATDTVA-----N 526
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      472 YVITREBYAAOEGEGSTLGPWTAQAQOLFVDMANVALRRLRLVEHSALAPDLSOQMN 531
Qy      527 MSSGPEPRFFKNLIASLPIADRAPIGKHFGDVLQAPKREYRVGEVVEYIEYGANPKNS 586
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      532 FQTVG-----VADDPYIKKSGDVLQOQRESRYRIGDQVTAFAVTGHPKH- 575
Qy      587 AEMQTHOTELTYEKY-EDSVADMOQIMYNDASMERPFYVHKILDLNSAITYWHPIDPAVP 645
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      576 -DLTRKTELVEVNIIGKDKQTPVTVAVDNDMDVOYRWERGIASAKATLSWISIPGETEP 634
Qy      646 GYIRIRYFGHNKROKELLPAVILAEIGISSEPEVNTT 682
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      635 GHYIIRHYNKAKNFWTOK---ISELGGSTRSEVILGT 668

```

RESULT 3

US-09-/739-90/-14b
Sequence 146, Application US/09739907
Patent No. US20010012889A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 36 Human Secreted Proteins
FILE REFERENCE: P2022P1
CURRENT APPLICATION NUMBER: US/09/739,907
CURRENT FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 09/348,457
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: 60/070,567
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,692
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,704
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,658
PRIOR FILING DATE: 1998-01-07
NUMBER OF SEQ ID NOS: 196
SOFTWARE: Patentlin Ver. 2.0

; LENGTH: 280
FIRST 2000

ORGANISM: *Homo sapiens*

US-09-739-907-146

Query match	24.58;	Score 879;	DB 10;	Length 280;
-------------	--------	------------	--------	-------------

Matches 163; Conservative 28; Mismatches 10; Indels 0; Gaps 0;

[illegible]

Db 241 KEMIVLKMVDLNGDGLISF 261

RESULT 4

US-09-739-907-76

Sequence 76, Application US/09739907

Patent No. US20010012889A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: 36 Human Secreted Proteins

FILE REFERENCE: P2022P1

CURRENT APPLICATION NUMBER: US/09/739,907

CURRENT FILING DATE: 2000-12-20

PRIOR APPLICATION NUMBER: 09/348,457

PRIOR FILING DATE: 1999-07-07

PRIOR APPLICATION NUMBER: 60/070,567

PRIOR FILING DATE: 1998-01-07

PRIOR APPLICATION NUMBER: 60/070,692

PRIOR FILING DATE: 1998-01-07

PRIOR APPLICATION NUMBER: 60/070,704

PRIOR FILING DATE: 1998-01-07

PRIOR APPLICATION NUMBER: 60/070,658

PRIOR FILING DATE: 1998-01-07

NUMBER OF SEQ ID NOS: 196

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 76

LENGTH: 318

TYPE: PRT

ORGANISM: Homo sapiens

US-09-739-907-76

Query Match 24.4%; Score 874; DB 10; Length 318;

Best Local Similarity 80.6%; Pred. No. 2,9e-68;

Matches 162; Conservative 28; Mismatches 11; Indels 0; Gaps 0;

QY 1 FSGYIGVGADCTGQVSDINLMGYNQGNARGLTRFSRAFIADPDGNSNMAFVS 60

Db 99 FSGYHIGVGADCTGQVADINLMGYSQNAQGLTRIXSRAFIMEPDGNSNRTVSVSI 158

QY 61 ELCMISQRLREVLAKRESKYGSIXRRDNVILSAIHSHSGPAGFOFTLYLAEGRSNR 120

Db 159 DIGWVSQRLREVLNIRKSGYSLYRRDNVILSGTHHSGPAGFOFTVFIASEGRSNQ 218

QY 121 TFOYIVSGIMKSIDIAHTNLKPGKIFINKGNVANVOINRSPSYLLNPOSERARYSSNTD 180

Db 219 TFOHMTGILKSIDIPHTNKKPGKIFINKGNVDSVOINRSPSYLLNPOSERARYSSNTD 278

QY 181 KEMIVLKLVDLNGEDGLISM 201

Db 279 KEMIVLKMVDLNGDGLISF 299

RESULT 5

US-10-242-056-59

Sequence 59, Application US/10242056

Patent No. US20030113323A1

GENERAL INFORMATION:

APPLICANT: Ensign, Jerald C

APPLICANT: Bowen, David J

APPLICANT: Petell, James

APPLICANT: Faily, Raymond

APPLICANT: Schoonover, Sue

APPLICANT: ffrench-Constant, Richard

APPLICANT: Orr, Gregory L

APPLICANT: Merlo, Donald J

APPLICANT: Roberts, Jean L

APPLICANT: Rochelleau, Thomas A

TITLE OF INVENTION: Insecticidal Protein Toxins from

NUMBER OF SEQUENCES: 88

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Dowelanco

STREET: 9330 Zionsville Road

CITY: Indianapolis

STATE: IN

COUNTRY: US

ZIP: 46268

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/242,056

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/063,615

FILING DATE: 18-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/395,497

FILING DATE: 28-FEB-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/007,255

FILING DATE: 06-NOV-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/608,423

FILING DATE: 28-FEB-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/705,484

FILING DATE: 28-AUG-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/743,699

FILING DATE: 06-NOV-1996

ATTORNEY/AGENT INFORMATION:

NAME: Borucki, Andrea T

REGISTRATION NUMBER: 33651

REFERENCE/DOCKET NUMBER: 50301E

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-337-4846

TELEFAX: 317-337-4847

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 1565 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-10-242-056-59

Query Match 3.7%; Score 131.5; DB 9; Length 1565;

Best Local Similarity 20.4%; Pred. No. 0.047;

Matches 141; Conservative 77; Mismatches 257; Indels 215; Gaps 32;

QY 37 TRLSRAFIADPDGNSNMAFVSVELCMISQRLLEVLKRLSKYGSIXRRDNVILSAIH 96

Db 656 TYTLSEADFTDPDKN-----YLQVCL-----NVWMDHYD 685

QY 97 THSGPAGFOQ-----YLYLASEGFSNRTFOYIVS-----GIMKSIDIAHTNLKPGKI 145

Db 686 RPSGKKGATSWKSWKFWNYVALQDSKAPDAIRLVSYRDSKRGVLQYLDFTSSL-PAKT 744

QY 146 FINKGNV-----ANVOINRSPSYLL--NPOSERARYSSNTDKEMIVLKLVDLNGEDLG 197

Db 745 RLNTFTVFTLLEKANLIGD-SLDYTLQADPSLEADLVTDGKSEPM-----DENGSN-G 796

QY 198 LISWFAIHPSVMNNSNHFVNSDNNGYAA-----YLFEDQ--KNKGYLPGQGPVAGFASSNL 252

Db 797 LYFWELFPHLPFLVATRPANEOQFSPAKSLHYTFDPAKKK-----PINAP-----AY 845

QY 253 GDVSPNIIIGPHCVGTGSCDNKSTCPNGSPGCMASGPGQDMFESTYIIIGRIIYOKA-- 310

Db 846 WNVKPLVWG-----NSDLSRHLDSDIDP-----DTQAYAH-----PVLYOKAVF 884

QY 311 -----KELYASASQEVYGP-----VLAH 329

Db 885 IAYVSNLIAQGDMMYROLTRDGLTQARVYVNLAAELIGPRPDVSLSSIMWPTQTLTLAAG 944

QY	330	QWVMTWDSVQNLNTHYKTKCPALGYSFPAAGTIDV--SGNITFQGTGEGPFEDTIRD	387
Db	945	QKAVLRDEHQIANSDFLALPAPGNNVSYLALADGYNENPLNVLMLS----	IHHDTIDA 999
QY	388	QLL-----GKPEIIVECQKPKPLL-----HSEELTIPHPWQDIYDVQIVYWS	433
Db	1000	RLYLNRHNLVDGKRPPLSLYLAAPYDVPALLAQRAQSSTLT-----NGVSGAMELTVP	1055
QY	434	LAIATAI-----GGEITLMSG-----RREPRAIKKEFALYGMKDM-----V	VIA 472
Db	1053	YRFSAEMPRAYSAVGTLTSPFCQNLTLSEBSERACQEBLAAQQLDMSSVAILTQQQALD	1112
QY	473	GLSNVYTHYITYEYQQRYEASTIYIGPHLSYITLFLFDLAKA-----IATDPAVAMS	528
Db	1113	GLADRLALALASQATAQ-QRHDHYTLT-QNNISSABELVMDTQTSASLSSISSTGVOTA	1170
QY	529	SGPEPPFFKNLIASLIPNIADRAPIGKHFQDVLPAKPEYRGEVEVYIFVGANKPNSAE	588
Db	1171	SG-----ALKYIPNIFGLAGGS-----RYEGVTEAIAIGLMAACAT	1208
QY	589	NOTHOTELVEKEDSVADQIMYNDASME	618
Db	1209	SVAERLATENTENRRRREMOIOTQOAOSE	1238

```

RESULT 6
US-09-739-907-150
: Sequence 150. Application US/09739907
: Patent No. US20010012889A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: 36 Human Secreted Proteins
: FILE REFERENCE: P2022p1
: CURRENT APPLICATION NUMBER: US/09/739,907
: CURRENT FILING DATE: 2000-12-20
: PRIOR APPLICATION NUMBER: 09/348,457
: PRIOR FILING DATE: 1999-07-07
: PRIOR APPLICATION NUMBER: 60/070,567
: PRIOR FILING DATE: 1998-01-07
: PRIOR APPLICATION NUMBER: 60/070,692
: PRIOR FILING DATE: 1998-01-07
: PRIOR APPLICATION NUMBER: 60/070,704
: PRIOR FILING DATE: 1998-01-07
: PRIOR APPLICATION NUMBER: 60/070,658
: PRIOR FILING DATE: 1998-01-07
: NUMBER OF SEQ ID NOS: 196
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 150
: LENGTH: 27
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-739-907-150

Query Match          3.5%; Score 127; DB 10; Length 27;
Best Local Similarity 88.9%; Pred. No. 0.00019;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0.

QY      76 RLESKYGLYRRDNIISAHTHSGPA 102
      II:|||||IIIIIIIIII  ||IIII
Db      1 RLQSKYGLYRRDNIISGTHHSGPA 27

RESULT 7
US-09-739-907-148
: Sequence 148. Application US/09739907
: Patent No. US20010012889A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: 36 Human Secreted Proteins
: FILE REFERENCE: P2022p1
: CURRENT APPLICATION NUMBER: US/09/739,907
: CURRENT FILING DATE: 2000-12-20

```

```

? PRIOR APPLICATION NUMBER: 09/348,457
? PRIOR FILING DATE: 1999-07-07
? PRIOR APPLICATION NUMBER: 60/070,567
? PRIOR FILING DATE: 1998-01-07
? PRIOR APPLICATION NUMBER: 60/070,692
? PRIOR FILING DATE: 1998-01-07
? PRIOR APPLICATION NUMBER: 60/070,704
? PRIOR FILING DATE: 1998-01-07
? PRIOR APPLICATION NUMBER: 60/070,658
? PRIOR FILING DATE: 1998-01-07
? NUMBER OF SEQ ID NOS: 196
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 148
? LENGTH: 26
? TYPE: PRT
? ORGANISM: Homo sapiens
? US-09-739-907-148

Query Match      3.5%  Score 125;  DB 10;  Length 26;
Best Local Similarity 84.6%;  Pred. No. 0.00027;
Matches 22;  Conservative 4;  Mismatches 0;  Indels 0;  Gaps 0

OY      10  RADCTGQVSDINIMGYGRKNGONANGL 35
|||||:|||||:|||||:|||||:|
Db      1  RADCTGQVADINIMGYGRKSGONAGI 26

```

```

oy          331 WV-NMTDVSVALNTHVYKTKPKALGY-----SFAAGTIDGVSLSLNTGGTEEDPFW 383
      :   :| |    ::::       +       :||| :| |   :| |   :| |
US-09-815-242-10572
Sequence 10572, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes In
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 10572
LENGTH: 766
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-815-242-10572

Query Match           3.3%; Score 116.5; DB 10; Length 766;
Best Local Similarity 21.4%; Pred.No. 0.32;
Matches 87; Conservative 55; Mismatches 139; Indels 125; Gaps 22
```


Db 205 WISSOTDKAIAK-----DGFKKIRPAKDYDNLVYLSALAKADMVGLNTRALTV--KYOD 258
 QY 384 TLRLDGLKPS-----EEIVECOKPKPILHS-----GELTIPHPW----- 419
 Db 259 NLSGRVOTPTLAVNROOEKTEIOFKPOTYTTISLTVSESEKAKMTOKPKYALKEROEAO 318
 QY 420 -----OPDIV-DVQ--ITVVGSLAIAIPGELTLM--SGRRPREAKKEFALYCKMD 466
 Db 319 LKVELSKOKGLVTDIOEKVTEN-----APLPYDLTEIOREANORFOFSAKKTLSL----- 369
 QY 467 MTVVIAGLSNRY--THYITTYEYOQOREASTIYGPTLSAYIOLFPDLAKALATOTVA 525
 Db 370 -----VQSLYEHKIVSY-----PRTDSKY-----LTTDTKG 396
 QY 526 NMSS-----GPE-PPEFKNLASLIPNIADRAPICKHFGDVLQPAKFEY----- 568
 Db 397 TMRERLAIADFSPEVAGYLKNGAVVQKQAFQNAKYTDHGLLPTREOFREYKLSNDEQ 456
 QY 569 RVGEVVEIVYG--ANPRNSAENQTHOTFLTVKEYEDSVADQIMYNDASMET----- 619
 Db 457 KIYQMIYORFLGLFAEP-----NOTKQKTVAFGKETFEVPHQNKVVAVAGKMTAEQPLS 511
 QY 620 RFYHHKGLIGLSNNTIYWHIPDTAVPGIYRIRFGHNRKQELLKPA 665
 Db 512 TVQOKGMVAPNFTINKELTSPKP--LTEGTLGKMEKHSIGTPA 556
 RESULT 9
 US-09-928-457-38
 ; Sequence 38, Application US/09928457
 ; Patent No. US20020164603A1
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: DNA, specific proteins and peptides
 ; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method
 ; NUMBER OF SEQUENCES: 99
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patentin Release #1.0, Version #1.30 (OEB)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/928,457
 ; FILING DATE: 2001-08-14
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/214,759
 ; FILING DATE: 199-12-10
 ; INFORMATION FOR SEQ ID NO: 38:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1981 acids aamin,s
 ; TYPE: acide aamin,
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FEATURE:
 ; NAME/KEY: Peptide
 ; LOCATION: 1..1981
 ; US-09-928-457-38
 Query Match 3.2%; Score 113; DB 9; Length 1901;
 Best Local Similarity 18.9%; Pred. No.2.9;
 Matches 131; Conservative 86; Mismatches 204; Indels 272; Gaps 32;
 QY 109 LYIIASGFSNRTQYIVSGIMKSIDIAHTNLK--PGKIFINKGVANVQINRSPSSYLLN 167
 Db 511 LALLADNITAKT-----TNLNTPGNLKYHTGKDLNINVDKDISASIIH 554
 QY 168 POSERARYSSNTDEKMLVLKLVLDNGEDGLISMFALHPVSMNNSNHHVYNSDNNGY----- 223
 Db 555 LKSDNAAHITGTSTTLASK--DM-GVEAGILL-----NVTNTNLPTNSGNLHIOAK 603
 QY 224 -----AAYLFROEKNKGYLPQGG-----PFVAGFASSNLGQVSPNLLGPHC- 264

Db 604 GNIOLRNTKLAAKALEFTTALOGNIVSGLHAVSADGHSVLANGNADFTGHNTLTAKAD 663
 QY 265 VNTGE-----SCNDKSTCPNGGSMOMAGSGQDM----- 295
 Db 664 VNAGSVGKRLKADNTNITSSG--DITLVGNGIQLDDGKORSINSKHSIKNNGNA 721
 QY 296 -----FESTH-----II 302
 Db 722 LKNNLVAKSGALNIHSDRALSTENTKLESTHNTHLNAQHERYTLNVDVAAHRHLSIT 781
 QY 303 GRITQAKELIYASASQEVTPVLAHQWVNMTDVSOLNTHYVTKCPALGYSEPAAGT 362
 Db 782 GSOIWOONDK--LPSANKLYANGVLA-----LNARYS-----QIADNT 816
 QY 363 IDGVSGLNIOTGT--EGDPFWPTLRDOLLGKSPSEELVEOQKPKPILH--SGELTI 415
 Db 817 TLKAGAILMTAGTALVYKGNINMSTVSKTLEDNAE-----LKLPLAGRLNIEAGSGTLTI 871
 QY 416 -PHPQPDIVQIVTVGSLAIAIPG-----ELTMSGRFREAIKKEFALYGM-- 464
 Db 872 EPANRISAHPTLSTIKTGKLLLSAKGNAGAPSAQVSSLEAKGNIRLVYTGETDLRGSKIT 931
 QY 465 --KDMTVV-----IAGLSNVYTHYITTYE-----EYQAOQYEAASTIYGPH 504
 Db 932 AGKNLVAVTTKGLNIEAVNNSFSNYPFOKAELNCKSELEQOIAOLKSSS-----PKS 987
 QY 505 -----LSATYOLFRLDKA-----IADTYANMS-----SGPEPP 534
 Db 988 KLIFLQERDRFLAFTQAIINKEVKKPKKEKYLQAKNISQNDILISAQIEISGSDIT 1047
 QY 535 FFKNL-----IASLIPNIADRAPICK--HFGDVL--OPAKPEYRWG 571
 Db 1048 ASKTLNLHAAGVLPKRAADSEAAIILDSITTOYELEGKTYKSHYDKALNKPRLTGRTG 1107
 QY 572 EVEVEY-----IFVGAN-----PKNSAENQTH 592
 Db 1108 VSIHAAALDDARIIIGASEIKAPSGSIDIKAH 1140
 RESULT 10
 US-10-066-551-1
 ; Sequence 1, Application US/10066551
 ; Publication No. US20030100071A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Apicella, M. A.
 ; APPLICANT: Gibson, J. L.
 ; APPLICANT: Edwards, J. W.
 ; APPLICANT: Scheffler, K.
 ; APPLICANT: Brown, E.
 ; TITLE OF INVENTION: Vaccine and compositions for the
 ; TITLE OF INVENTION: Prevention and treatment of Neisserial infections
 ; FILE REFERENCE: 875.045051
 ; CURRENT APPLICATION NUMBER: US/10/066,551
 ; CURRENT FILING DATE: 2002-01-31
 ; PRIOR APPLICATION NUMBER: US 60/344,452
 ; PRIOR FILING DATE: 2001-10-23
 ; PRIOR APPLICATION NUMBER: US 60/310,356
 ; PRIOR FILING DATE: 2001-08-06
 ; PRIOR APPLICATION NUMBER: US 60/266,070
 ; PRIOR FILING DATE: 2001-01-31
 ; NUMBER OF SEQ ID NOS: 12
 ; SEQ ID NO 1
 ; LENGTH: 2015
 ; TYPE: PRT
 ; ORGANISM: Neisseria gonorrhoeae
 ; US-10-066-551-1
 Query Match 3.2%; Score 113; DB 9; Length 2015;
 Best Local Similarity 18.9%; Pred. No.2.9;
 Matches 131; Conservative 86; Mismatches 204; Indels 272; Gaps 32;
 QY 109 LYIIASGFSNRTQYIVSGIMKSIDIAHTNLK--PGKIFINKGVANVQINRSPSSYLLN 167

Db 511 LALLADNDITAKT-----TNLTPTGNTLVYHTGKDLNLVNDKDLASASIH 554
QY 168 POSERARSSNTDKEMLVKLVDNGEDGLISMFAIHPSVMNNSNHFFVSDMWG-----223
Db 555 LKSDMAAHTGTSKITLASK--DM-GVEAGL-----NNTNTLRTNSGMLHIOAK 603
QY 224 -----AAYLEFOEKNKGYLPGQ-----PVAFASSNLGDVSPNIIQPHC- 264
Db 604 GNQLRNTKLMAKALLETALQNIIVSDGLHVASADQHSLLANGNADFGHNTLAKAD 663
QY 265 VNTGE-----SCDNKSTCPNPGSPMCASGPGOM-----295
Db 664 VVAGSVGKRLADNTNITSSG--DITLVAGNGIOLGDKORNSINGKHISIKNGGNA 721
QY 296 -----FESTH-----II 302
Db 722 DLKLNVAHAKSGALNHSDBALSTENTKLESTHNTLMOHERVTLNOVDVAHRLHST 781
QY 303 GRITVOKAKELYASASQEVTPVLAHOMVMTDVSVOLNATHTVTKCPALGYSFAGT 362
Db 782 GSGIQNDK--LPSAKKLVANGYLA-----LNARYS-----QIADNT 816
QY 363 IGVSGSLNTGQTT-----EGDPFMDTLRDOLGKPSSEIVECOKPKPILH-----SGELT 415
Db 817 TLKAGINLTAGTALVKRCNINMTSTVSTKLEEDNAE-----LKPLAGRLNIEAGSGTLTI 871
QY 416 -PHMPQDVLVDQIYTVGSLATAIAPG-----ELTMSGRFRRAIKKEPALYGM-----464
Db 872 EPANRISANTDLSTKTKGKLLLSAKGNGAPSAOVSSLEAKNITLVGETDLKRSKIT 931
QY 465 --KDMTV-----IAGLSNVTHTYTYE-----EYQARFEASTIGPHT 504
Db 932 ACKNLVAVTTGKLNIEAVNNSPSTFPOKAELNQSKELEQOIAQJLKKSS-----PKS 987
QY 505 -----LSATIOLEFRLAKA-----IATDVANMS-----SGPEPP 554
Db 988 KLIPTLOERDLAFYIOAINKEVKKKPKKREYIOAKLSAONIDISAGIEISGSDIT 1047
QY 535 FFKNL-----IASLIPNIADRAPICK-----HFGDVL--QPAKPEYRVG 571
Db 1048 ASKKLNLHAAGVLPKAADEMAAILDITDYOIEIKPTTKSHYDAALNKPRLGRIG 1107
QY 572 EYEV-----IEVGAN-----PKNSAENQTH 592
Db 1108 VSIHAAALDARIIIGASEIKAPSGSIDIKAH 1140

RESULT 11
US-09-338-723A-4
; Sequence 4, Application US/09338723A
; Patent No. US20020019038A1
; GENERAL INFORMATION:
; APPLICANT: Huang, Wang
; TITLE OF INVENTION: Phenol Oxidizing Enzymes
; FILE REFERENCE: GC361-2
; CURRENT APPLICATION NUMBER: US/09/338, 723A
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 09/220, 871
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 627
; TYPE: PRF
; ORGANISM: Bipolaris spicifera
US-09-338-723A-4

Query Match 3.1%; Score 111.5; DB 10; Length 627;
Best Local Similarity 20.8%; Pred. No. 0.63;
Matches 129; Conservative 76; Mismatches 230; Indels 185; Gaps 35;
145 IFKNGVNAVNOINSPSSYLLN-POSERARYS--NTDKEMLVKLVDNGEDGLI-- 199

Db 13 VSIAGK-IYGVALLSEBPAPFVNDTPDEKKAALASIVEDPADVNNLKMOSPEYELIR 71
QY 200 SFAIHPVSMNN--SNHFPVSDMWGYAAYL--FEOEKNKGYLPGQGPVAFGASSNLGDV 255
Db 72 QPLPPIPAKEPNKLTIPVINKELIYETIKPPTQVYYSLRAR--LVGIGISPG-- 126
QY 256 SPNIIQ-----HCVNTGESCDNKSSTCPNPGSPMCASGPGDMFESTHIIGRITIQ 308
Db 127 -PTIIVPRGEAVVRING--DRESSIHLHSPSRAPFDDGADM-----IMG 173
QY 309 KAKELYASASQEVTPVLAHOMVMTDVSVOLNATHTVTKCPALGYSFAGTIDVSG 368
Db 174 EYKDYYPNNQ-----AARFLYHD-----HAMHV-----AENAYFGAG 209
QY 369 LNTGCTGSDPPMDTLRQGLGKPS-----EEIVECOKPKPILHSGELIIPHPQDVI 424
Db 210 AVLIT-----DEPEDA-----LGLPBGYGYKD-----PLVLSKYYNADGTLKTSVG 252
QY 425 DVQIYTVGSLATAIPEGETLMSGRFR-----EAIKKEFALYGMK--DMTV-----VIA 472
Db 253 EDSKSVGDIHVGQCPWPLNVPRKYLRLFLNAAVSRRFALYFKQDNTATRLPROVLA 312
QY 473 GLSNVTHTYTYEY--QAQRYEASTI--YGPHTLSAYIOLFRDLAKA-----518
Db 313 SDAGLLTHVQTSMDVAAAEYEIFDEPAPYAGOTLD-----LRNFAKANGIGTDDBYA 367
QY 519 -----IATDVANMSGSPPEPFKNLILPINADRAPICKHFDVLOPARPEYR 569
Db 368 NTDKVRFPNHSSTQVVDNSVPEQ-----LSQIOFPADKDDIDHFR--FHRINGEWR 418
QY 570 VGEV-----VEVIEVGANPK-----NSAENQTHQ-----TELVEKEDSVADMOI 610
Db 419 INIGRADVENVRLAVPPTGVELMELENSGGSWSPHIVHLVDFVARYGEGCRGYM 478
QY 611 M-----NDASMETRYMKIGILGSLNATIIYHIPTAIPGII-----648
Db 479 PEYAGLKVVWVGR--HETVL-----VEAHY--APWDGYMFCHNLIHEDODMAA 527
QY 649 ----RIRYFGHNRKOELEK 664
Db 528 FDTKRLQNGCYNETTDFHPD 547

RESULT 12
US-10-080-210-4
; Sequence 4, Application US/10080210
; Patent No. US20020142423A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Huang
; TITLE OF INVENTION: Phenol Oxidizing Enzymes
; FILE REFERENCE: GC561-3
; CURRENT APPLICATION NUMBER: US/10/080, 210
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/220, 871
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 09/338, 723
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 627
; TYPE: PRF
; ORGANISM: Bipolaris spicifera
US-10-080-210-4

Query Match 3.1%; Score 111.5; DB 12; Length 627;
Best Local Similarity 20.8%; Pred. No. 0.63;
Matches 129; Conservative 76; Mismatches 230; Indels 185; Gaps 35;
145 IFKNGVNAVNOINSPSSYLLN-POSERARYS--NTDKEMLVKLVDNGEDGLI-- 199

Db 13 VSANG-ITGVALSRAKFDVNTPEDEKALASIVEDDADVYNNMLKDMCSPDEPLIFR 71
QY 200 SWEAIHPVSMNN--SNHFVNSDNMGYAYL--PFOEKNKGYLPQGGFVAGFASNSLGDV 235
Db 72 GELPFPKAPKRLNPTNKEIWEIYELVIFPQOYPPSLRPAR--LVGYDISPG-- 126
QY 256 SPNLLGP-----HCVTGESCNDKSTCPNGSPCMASGPRODMEFHILIGRIIYO 308
Db 127 -PITIVPGRTEAVNRING--DRESSIHLGSPSRAPFGMDDM-----IMKG 173
QY 309 KAEIYASASQEVTPVLAHOMVNMTOVSQNLNPTNKTCPALGYSFNAGITDYSG 368
Db 174 EYKDYYPNNQ-----AARFLMYHD-----HAMHY-----AENAYGQAG 209
QY 369 LNTGCTEGDPPMDLTDLLGKPS--EETVECCQKPKPLLSGELTTPHMQPIIV 424
Db 210 AYLTIT-----DPAEDA-----LGLPSGYGKYDI-----PLVLSKYYNADGTLKTSYG 252
QY 425 DVOIVTVGSLAIAIPGELTTMGRPR-----EAIKKEPALTGMK--DMTV-----VIA 472
Db 253 EDKSWMGDIHVNGQWMPFLNVEPRKRLFLNNAVSRNFALYFVQODMTATRLPFOVIA 312
QY 473 GLSNVYHYITTYEY--QAORYEASTI--YGPHTLSAVIQLFRLAKA----- 518
Db 313 SDAGLTHPVSOTSMYVAAAEYELVDFAPYAGQTLID-----LRNFAKANGIGDDDYA 367
QY 519 -----IATDVANNSGCEPPEFFKNLILASIPRIADRAPIGKHIVLOPARREY 569
Db 368 NTDKVRFHVSQTFVVDNSVVEQ-----LSQIOPRACKTDIDHFR--FRTNGEMR 418
QY 570 VGEV-----VEVIFVGANKP-----NSAENQTHQ-----TFLTEKYEVSADMOI 610
Db 419 INGIGADVENVKILAKVPRGIVELMELNSSGMSHPRIHVILVDFRVARAYGDEGTGVM 478
QY 611 MY-----NDASWETRPYMKGLIGLSNATYIWHIPDAYPGY----- 648
Db 479 PYEAGLKDYVWLGK--HETVL-----VEAHY--APWDGYVMHCHNLHEDODMMAA 527
QY 649 -----RTRYGHNKROELLKP 664
Db 528 FDVTKLQNGFNETTDEHDP 547

RESULT 13
US-10-303-685-14
; Sequence 14, Application US/10303685
; Publication No. US20030100005A1
; GENERAL INFORMATION:
; APPLICANT: Exelixis, Inc.
; TITLE OF INVENTION: CRBS AS MODIFIERS OF BRANCHING MORPHOGENESIS AND METHODS OF USE
; FILE REFERENCE: EX02-125C
; CURRENT APPLICATION NUMBER: US/10/303,685
; CURRENT FILING DATE: 2002-11-25
; PRIOR APPLICATION NUMBER: 60/333,388
; PRIOR FILING DATE: 2001-11-26
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1376
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-303-685-14

Query Match 3.0%; Score 106; DB 9; Length 1376;
Best Local Similarity 22.7%; Pred. No. 6.6;
Matches 77; Conservative 41; Mismatches 97; Indels 124; Gaps 22;

QY 73 VLKRLSKYGLYRDNVILSAIHHSQAGFFOYTLVLAEGESNRTFOYVIGIMKS 132
Db 725 VIFLDESYSG-----DTISLSMVRTLQPSGL-----LLALE--NSTYQYI----- 763
QY 133 IDIAHNTLKPCKIFINKGNVANYQINRSPSSYLLNPOSERARYSSNTDKEMVLYKLVDLN 192
Db 528 FDVTKLQNGFNETTDEHDP 547

Db 764 -----RWVLEGRILA-----MLTPNSPK-----LVYKFV--LN 789
QY 193 GEDILISWFAIHPVSMNNSHFNVDNMGY--AAVLEDEKRNK-----GYLPQO-----G 241
Db 790 DGNVHLIS-LKIKPKYK--ELYOSSQNLGFIASSTWIKIEKGVYITIGLPDQOETELNG 845
QY 242 PVVAG-----FASNLGDVSPN-----ILGPHCVNTGESCNDKS-----TCPNGGPRSMC 286
Db 846 GFPGKCIODVRLNQNLT--EFPNPTNNAASLNPVLYVNTQOCAGDSCSKSPCHNG-- 900
QY 287 MASGPGDMFESTHIIIGRIIYOKAEIYASASQEVTPVLAHOMVNMTOVSQNLNPTNKTCPALGYSFNAGITDYSG 368
Db 901 -----VCHSRWDDFSCSCPALTSKACEEYQWCGFSP-----CPHG 936
QY 347 VTKCKPAL--GYSFPACTI--DGVSGL-----NITGCTT 376
Db 937 AQ--CQPVLOGFECIANAVFNGSGOILFRSNGNITRELT 974

RESULT 14
US-10-303-685-15
; Sequence 15, Application US/10303685
; Publication No. US20030100005A1
; GENERAL INFORMATION:
; APPLICANT: Exelixis, Inc.
; TITLE OF INVENTION: CRBS AS MODIFIERS OF BRANCHING MORPHOGENESIS AND METHODS OF USE
; FILE REFERENCE: EX02-125C
; CURRENT APPLICATION NUMBER: US/10/303,685
; CURRENT FILING DATE: 2002-11-25
; PRIOR APPLICATION NUMBER: 60/333,388
; PRIOR FILING DATE: 2001-11-26
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 1406
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-303-685-15

Query Match 3.0%; Score 106; DB 9; Length 1406;
Best Local Similarity 22.7%; Pred. No. 6.9;
Matches 77; Conservative 41; Mismatches 97; Indels 124; Gaps 22;

QY 73 VLKRLSKYGLYRDNVILSAIHHSQAGFFOYTLVLAEGESNRTFOYVIGIMKS 132
Db 725 VIFLDESYSG-----DTISLSMVRTLQPSGL-----LLALE--NSTYQYI----- 763
QY 133 IDIAHNTLKPCKIFINKGNVANYQINRSPSSYLLNPOSERARYSSNTDKEMVLYKLVDLN 192
Db 764 -----RWVLEGRILA-----MLTPNSPK-----LVYKFV--LN 789
QY 193 GEDILISWFAIHPVSMNNSHFNVDNMGY--AAVLEDEKRNK-----GYLPQO-----G 241
Db 790 DGNVHLIS-LKIKPKYK--ELYOSSQNLGFIASSTWIKIEKGVYITIGLPDQOETELNG 845
QY 242 PVVAG-----FASNLGDVSPN-----ILGPHCVNTGESCNDKS-----TCPNGGPRSMC 286
Db 846 GFPGKCIODVRLNQNLT--EFPNPTNNAASLNPVLYVNTQOCAGDSCSKSPCHNG-- 900
QY 287 MASGPGDMFESTHIIIGRIIYOKAEIYASASQEVTPVLAHOMVNMTOVSQNLNPTNKTCPALGYSFNAGITDYSG 368
Db 901 -----VCHSRWDDFSCSCPALTSKACEEYQWCGFSP-----CPHG 936
QY 347 VTKCKPAL--GYSFPACTI--DGVSGL-----NITGCTT 376
Db 937 AQ--CQPVLOGFECIANAVFNGSGOILFRSNGNITRELT 974

RESULT 15
US-10-165-049-2
; Sequence 2, Application US/10165049
; Publication No. US20020192724A1
; GENERAL INFORMATION:

```

APPLICANT: Brenner, Michael
APPLICANT: Cepek, Karyn
TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
FILE REFERENCE: L00560/70010ERP
CURRENT FILING DATE: 2002-06-07
PRIOR APPLICATION NUMBER: USSN 08/237,919
PRIOR FILING DATE: 1994-05-03
PRIOR APPLICATION NUMBER: PCT/US 95/05518
PRIOR FILING DATE: 1995-05-03
PRIOR APPLICATION NUMBER: USSN 08/732,429
PRIOR FILING DATE: 1996-11-01
PRIOR APPLICATION NUMBER: USSN 09/798,267
PRIOR FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 878
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (800)..(808)
OTHER INFORMATION: HAV tripeptide
FEATURE:
NAME/KEY: misc_feature
LOCATION: (2225)..(2295)
OTHER INFORMATION: transmembrane
US-10-165-049-2

```

Query Match 2.9% Score 104; DB 9; Length 878;

Best Local Similarity 18.6% Pred. No. 4.9; Mismatches 130; Conservative 94; Indels 220; Gaps 32;

```

47 ADPGSNRAEYVSVLCMISORLELV-----LKR-----LESKYG 82
41 APPEKRPRLGRVNFEDCTGRRTAIFLPKVGTDGVTYKRLRPHNPTDPLGLRWD 100
83 SLXRR--DNVILSAHTHSQPAFGFYLYLASGFSNRTFOYIVSGIMKSIDIAHTNL 140
101 STYKFEKSTKVTLNTGVGHHRRP-----PHQASVSGIOAEL-LTFPNS 141
141 KPGKIFINKGNVANYQINKSPSSYLLN---POSERARY-----SSNTDKEMLVKLV 189
142 SPG-----LRQKRDWVLPISCPENEGKPEPKLVQIKSNKDEKGVYSI 188
190 DLNGEDLGLISWFAI-----HPVSMNNSNHFVNSDNMGYA 224
189 TGQGADTPPVGVFIIRETGWLKVTPELDRERIATYTLFSAVSSNGN---AVEDPMETL 245
225 AYLFQOEKMK-----GYLPQGPFFVAGFASNLGDVSPNLPCHVCNTGES 270
246 ITVTQDNQNKPEFTQEVKRGSMEGALPC-----TSMEV-----TATD 284
271 CDNDKSTCPNGSPSCMASG---PQDMF---ESTHIIIGRIIYQAKELVASASQEVGT 323
265 ADDDVNTY-NAIAIATYILISOPELDPKMTINRNTGVISVTTGLDRESFPTYLVOA 343
324 PYLAHQVNMVDVSVQNLATHTVTKCKALGYSPAAGTIDG-----VSGLINTQ 373
344 ADLOG-EGLSSTATAV---ITVTDNDNPI---FNPTTYKGQVPENANVVITLKVTD 396
374 GTTEGDPFMDTLRDOLGKPSSEIVECQK-----KPLLHSGELTIIPHPQDPDI 423
397 ADAPNTPAMEAVY-TILNDGCGFYVTTNPVNDGILKTAAGIDPEAKQOYILHVAVTNV 455
424 VDOQIYTVGSLAIAIPGELTTMSGRRF-----REAIKKEFALYGMKDMTVYIAGLSNVY 478
456 VPEFVSLTSTATVTV-DVLVDNNEGTFVPEPEKRVSEDFGV--CGEITSYTAQEDPTE 512
479 THYITTYEYQORREASTIIGPHT--LSAVIQLFRD-----LAKAATDTPVAN 526

```

```

DB 513 MEOKITVRIMWDRNMLE--INPDGAISTRALREDEFEHVNSTYTAIIITDNGSP 569
QY 527 MSSGPEPFFKNLIALIPNIADRAPIGKHGVDLQPAKPEYRVGEVEVIFVGANKR-- 584
DB 570 VATG-----TGTLLILSDVNDNAP-----PEPR-----IFFCERNKRPQ 606
QY 585 ---NSAENQTHQFLTVKEYEDSVADQIYNDAWET 619
DB 607 VINIHADLPNPTSPEFTALTHGRVPMWTIQNDPTQES 645

```

Search completed: July 3, 2003, 12:56:14
Job time: 57 secs